PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, 1/36, A61K 35/74, C07K 14/255, C12N 15/74, 1/21, C12Q 1/68, C07K 16/12, A61K 39/112, 48/00, G01N 33/53

A2

(11) International Publication Number: **WO 99/45120**

(43) International Publication Date: 10 September 1999 (10.09.99)

(21) International Application Number: PCT/GB99/00675

(22) International Filing Date: 8 March 1999 (08.03.99)

(30) Priority Data:
9804809.3 6 March 1998 (06.03.98) GB

(71) Applicant *(for all designated States except US)*: BIOTECHNOLOGY AND BIOLOGICAL SCIENCES RESEARCH COUNCIL [GB/GB]; Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1UH (GB).

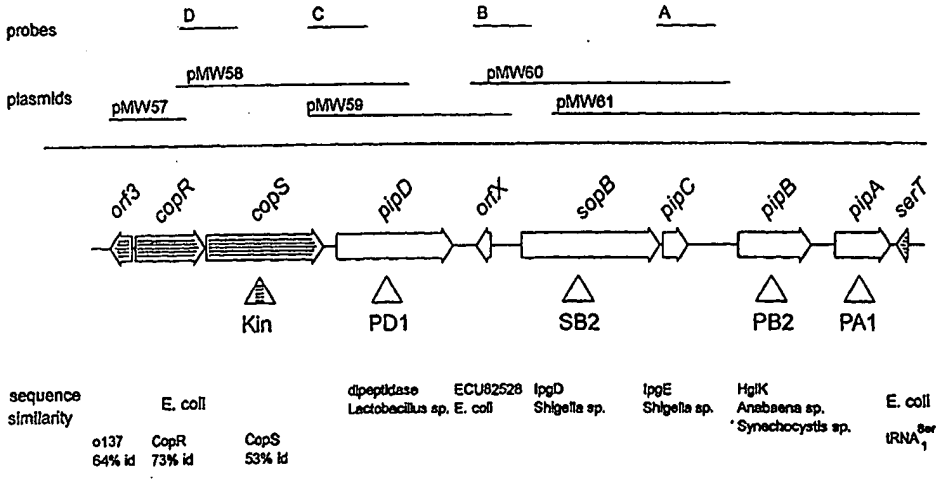
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: WALLIS, Timothy, Stephen [GB/GB]; Grange Cottage, Yattendon, Berks RG18 0UE (GB). GALYOV, Edouard Eugenievich [RU/GB]; 30 Humber Close, Didcot, Oxon OX11 7RU (GB). WOOD, Michael, William [GB/GB]; Roden House, Wallingford Road, Compton, Berks RG20 6QR (GB).

(74) Agents: O'BRIEN, Caroline, J. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States: GB, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: ATTENUATION OF BACTERIA: MATERIALS AND METHODS RELATING THERETO



(57) Abstract

Attenuated microorganisms are prepared by mutating one or more genes selected from *sopA*, *sopD*, *sipA*, or a gene from within the pathogenicity island SPI-5, when the microorganism is *Salmonella*, or one or more homologous genes which correspond to said *Salmonella* genes, when the microorganism is another enteric bacteria. In comparison with the wild-type strains, the attenuated microorganisms have a reduced enteropathogenicity but substantially the same invasiveness. The attenuated microorganisms can be administered as vaccines against infection by enteric bacteria. Polypeptides encoded by *sopA*, *pipD*, *orfX*, *pipC*, *pipB* and *pipA* genes, variants thereof, and nucleic acids encoding those polypeptides, are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

ATTENUATION OF BACTERIA: MATERIALS AND METHODS RELATING THERETO

Field of the Invention

The present invention concerns attenuation of bacteria and materials and methods relating thereto. In particular the present invention concerns vaccines for *Salmonella* and materials and methods relation thereto.

Background to the invention

Salmonella comprises a genus of aerobic to facultatively anaerobic bacteria which cause a variety of effects in mammals: diarrhoea and/or abortion in some species of domestic livestock and enteric fever/food poisoning in humans. Some *Salmonella* serotypes cause predominantly enteric effects, others predominantly systemic effects.

Salmonella infection in young and elderly people can be particularly debilitating or even fatal. There remains a need for effective vaccines which are safe to use.

A characteristic feature of non-typhoid *Salmonella* infections in animals is an intense intestinal inflammatory and secretory response (acute inflammatory cell influx, fluid secretion) typical of enteritis. Although the precise mechanism by which non-typhoid salmonellae disrupt normal intestinal function and induce enteritis is not known, recent findings suggest that interactions between intestinal epithelial cells and the pathogen play a key role in mediating the inflammatory response (Eckman et al., 1993; Galyov et al., 1997;

McCormick et al., 1993, 1995a,b). It has been shown that enteropathogenic *Salmonella* strains are able to induce intact intestinal epithelia to recruit sub-epithelial neutrophils (PMNs) (McCormick et al., 1995a,b). This *Salmonella* mediated transepithelial signalling to PMNs requires adhesion of *Salmonella* to the epithelial apical membrane and is dependent on protein synthesis in both bacteria and eukaryotic cells (McCormick et al., 1993, 1995a). *Salmonella* mutants defective in the function of a type III secretion system encoded by the *inv/spa/prg* chromosomal loci were shown to be unable to elicit transepithelial signalling to PMNs (McCormick et al., 1995a,b). In addition, the Inv/Spa secretion system of *Salmonella* is required for the ability of this pathogen to invade epithelial cells (for a review see Galán, 1996).

The genes encoding the structural components of the Inv/Spa secretion system, a number of secreted proteins and associated regulatory proteins are clustered at 63 centisome on the *S.typhimurium* chromosome. The analysis of these loci revealed that they constitute a pathogenicity island (SPI-1), a large inserted DNA fragment which confers upon the host bacterium virulence associated functions (Mills et al., 1995). A second pathogenicity island (SPI-2), was found to encode components of a second type III secretion system in *Salmonella* (Ochman et al., 1996; Shea et al., 1996). SPI-2 is located at 30 centisome (Shea et al., 1996). The genes located on the SPI-2 appear to be important for the development of systemic infection (Shea et al., 1996). Finally, a *Salmonella* virulence-associated locus at 82 centisome was recently identified as a novel

pathogenicity island, SPI-3. SPI-3 encodes mtgCB genes essential for intramacrophage survival of the pathogen (Blanc-Potard and Groisman, 1997).

A number of proteins secreted by the Inv/Spa secretion system have been characterised. Among them, the Sip proteins (*Salmonella* invasion proteins), are essential for invasion (for a review see Galán, 1996) and play an important role in the translocation of Sop proteins (*Salmonella* outer proteins) into the target cell (Galyov et al., 1997; Wood et al., 1996). The Sop proteins appear to have effector functions. The present inventors have recently characterised the SopB protein and shown that the translocation and intracellular activities of SopB (and possibly other Sop effector proteins) affect cellular responses leading to the influx of neutrophils into the intestinal epithelium and the induction of fluid secretion (Galyov et al., 1997). The *sip* genes are constituents of SPI-1. Herein the inventors report that *sopB* is located on a large DNA fragment unique to the *Salmonella* chromosome and representing a new pathogenicity island, SPI-5 (this pathogenicity island has been previously referred to as SPI-4 by the present applicants eg in their UK Patent Application No. 9804809.3 filed on 6 March 1998) inserted adjacent to the $tRNA_1^{Ser}$ (*serT*) gene. The structure of SPI-5 and its boundaries in relation to the sequence of the *E.coli* K-12 genome is described herein and data is provided indicating that SPI-5 encoded proteins are involved in the enteropathogenicity of *Salmonella*. The SPI-5 is conserved in Salmonellae and maps at approximately 20 centisome of *S.typhimurium* chromosome. Sequence analysis reveals that the *sopB*-containing *Salmonella*-specific DNA

fragment is flanked by DNA sequences sharing a significant sequence similarity with genes found in *E.coli* K-12, $tRNA_1^{Ser}$ (*serT*) on one side and *copS/copR* on the other. Thus, this *Salmonella*-specific DNA fragment has features characteristic of "pathogenicity islands".

Pathogenicity islands are horizontally acquired clusters of genes, often inserted at tRNA loci and conferring upon the host bacterium certain virulence features. The present inventors have surprisingly identified a novel pathogenicity island, SPI-5, in *S.dublin* that mediates enteropathogenesis. Features of SPI-5 are as follows: SPI-5 encoded genes are necessary for the enteropathogenic phenotype; SPI-5 is conserved in *Salmonella*, but not present in the other enteric bacteria tested in this study; SPI-5 is located immediately downstream of the *serT* gene encoding $tRNA_1^{Ser}$. SPI-5 is also inserted between the *serT* and *copR/copS* genes. The *copS/copR* genes map at 46 centisome on the *E.coli* K-12 chromosome , whereas the *serT* maps at 23 centisome. The acquisition of the SPI-5 may be concomitant with a rearrangement of the *serT*, *copS/copR*, ORF and possibly some other genes in the chromosome of *S.dublin* 2229 compared to that of *E.coli* K-12. Alternatively, this region of the *S.dublin* chromosome may represent a "hot spot" for recombination events with rearrangements of this area of the *S.dublin* chromosome occurring secondary to the acquisition of SPI-5 insertion-recombination events.

In addition to the previously characterised *sopB* gene, SPI-5 contains other genes associated with enteropathogenicity of *S.dublin*. The putative products

of two of these genes, *pipB* and *pipD*, have structural similarities to proteins from other bacterial species. The observed sequence similarities are suggestive of functions of these SPI-5 encoded proteins. PipB is similar to Hg1K protein from *Anabaena* and *Synechocystis* spp. Since Hg1K appears to be required for localisation of heterocyst specific glycolipids (Black et al., 1995), it is suggested that PipB has a role in glycolipid biogenesis. PipD is structurally similar to dipeptidases from *Lactobacillus* spp and is suggested to be a peptidase of *Salmonella* having its target on the surface or inside the host cells.

SPI-5 was sequenced and, in addition to *sopB*, was found to contain five novel genes which the inventors have named *pipA*, *pipB*, *pipC*, *pipD* (pathogenicity island encoded proteins) and *orfX*. Each of the *pipA*, *pipB and pipD* genes was mutated and the enteropathogenicity of the mutant strains assessed in bovine ileal ligated loops. The results of the study indicate that the *pipA*, *pipB and pipD* genes contribute to the enteropathogenicity of *S.dublin*.

The present inventors have also identified and characterised a previously unknown gene which they call *sopA*. Details relating to this new gene *sopA* are provided herein. In particular the present application provides a demonstration that *sopA* has translocation and intracellular activities which affect cellular responses leading to the influx of neutrophils into the intestinal epithelium and the induction of fluid secretion (akin to *sopB*). Thus an assessment of the enteropathogenicity of *sopA* mutant *S.dublin* shows that *sopA* contributes to

enteropathogenicity.

The *sopA* gene locates close to the SPI-1 pathogenicity island which embraces the structural components of the Inv/Spa secretion system. The present application provides sequence information for *sopA* and instructions for its cloning and the construction of *sopA* mutants.

Summary of the invention

The present invention provides attenuated strains of *Salmonella* and attenuated strains of other enteric bacteria which have a homologue of a *Salmonella* protein/gene discussed herein; particularly provided are attenuated strains of *Salmonella, Shigella* and *E.coli*. The present invention also provides prophylactic and therapeutic medicaments which contain as an effective agent one or more attenuated strains of *Salmonella* and other enteric bacteria (eg *Shigella* and *E.coli*) as hereby provided. The present invention also provides use of one or more attenuated strains of *Salmonella* and other enteric bacteria (eg *Shigella* and *E.coli*) as hereby provided in the preparation of a medicament comprising the attenuated strain(s) as an effective agent; the medicament being for the treatment of, or immunisation against, infection by *Salmonella* and other enteric bacteria (eg *Shigella* and *E.coli*) and/or for the treatment of, or immunisation against, infection with a pathogen different to the mutant pathogen employed for immunisation, nucleotide sequences for antigens of said different pathogen being carried and expressible by attenuated strains as hereby provided. On the basis of sequence similarity, an attenuated strain of *Salmonella*

as hereby provided may also be useful in relation to the treatment of other enteric bacteria (eg *Shigella* or *E.coli*). Similarly an attenuated strain of other enteric bacteria (eg *Shigella* or *E.coli*) as hereby provided may also be useful in relation to the treatment of and/or immunisation against *Salmonella* and other enteric bacteria (eg *Shigella* and *E.coli*).

As stated above, the present invention provides attenuated strains of *Salmonella* and attenuated strains of some other bacteria, particularly enteric bacteria such as *Shigella* and *E.coli* which carry a homologue of a *Salmonella* protein/gene discussed herein.

The homologous versions in such other enteric bacteria will provide sites for achieving attenuation in accordance with the present invention.

Homologues of a *Salmonella* protein/gene discussed herein in other enteric bacteria having at least 20% overall sequence identity may be determined in accordance with standard techniques and by comparative analysis of sequences available on public data bases.

The nature of the attenuations is more fully described below. Briefly attenuation is effected by alterations in one or more genes which contribute to the enteropathogenicity of an enteric bacteria eg *Salmonella* species so as to reduce the enteropathogenicity of the altered form of the bacteria eg *Salmonella* species in comparison to its wild type with or without affecting invasiveness. Reduction in enteropathogenicity may be determined by measuring intestinal secretory response

and/or intestinal inflammatory response (see later examples and Fig 4 and Fig 7). In particular the attenuation in *Salmonella* may be effected by one or more alterations in one or more genes selected from *sopD*, *sipA*, *sopA* or a gene from within the pathogenicity island SPI-5 which comprises 6 genes including *sopB* and is flanked by the genes *copS/copR* and *serT* and wherein where there is an attenuation in the gene *sopB* of SPI-5 another said gene comprises an alteration. Attenuation in other enteric bacteria eg *Shigella* and *E.coli* may be affected by alteration of genes in the subject bacteria equivalent to the *Salmonella* genes discussed herein.

Genes from within the SPI-5 pathogenicity island suitable for alteration may be selected from *pipD*, *pipB*, *pipA*. The attenuation may be effected by an alteration in the gene *sopD*. (*SopD* was first sequenced by Ostrowski et al., see J.Biol. Chem., 1989 264(26) p15726-37, designated orf4 and identified as lying adjacent to SPI-1. No function was ascribed to *sopD*). The attenuation may be effected by an alteration in the gene *sipA*. (*SipA* was first described by Kaniga et al., see J.Bacteriol., 1995 177 p7078-7085 and lies within SPI-1. No function was ascribed to *sipA*). The attenuation may be effected by an alteration in the gene *sopA* as provided herein. The attenuation may be effected by alterations in *sopB* and another gene as mentioned above. In particular the attenuation may be effected by alterations in *sopD* and *sopB*. Alterations in both *sopD* and *sopB* is particularly and surprisingly advantageous in the context of attenuation.

The means of achieving attenuation as disclosed herein eg

by alteration of one or more of *sipA, sopA, pipD, pipB, pipA, sopD, sopB* reduces enteropathogenicity. A given strain of *Salmonella* may also be attenuated in accordance with known techniques to reduce systemic pathogenicity. For example, the attenuation of systemic pathogenesis may be achieved by curing of virulence plasmid (Wallis, T.S., et al., 1995) or by altering one or more genes involved in the aromatic amino acid biosynthetic pathway (Jones, P.J., et al., 1991 Vaccine 9, 29-34).

A medicament comprising as an effective agent an attenuated strain of *Salmonella* or other enteric bacteria as discussed above may comprise, in addition, one or more physiologically acceptable carriers such as excipients, buffers, stabilisers, adjuvants or other materials known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material will depend on the route of administration. Examples of techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 16th edition, Osol, A. (ed), 1980. Such a medicament may comprise one or more different effective agents eg the medicament may be a combination vaccine for use in relation to a variety of pathogens.

Thus a said medicament may be administered alone or in combination with other treatments, either simultaneously or sequentially, dependent upon the condition to be treated.

The medicaments of the present invention are preferably

given to an individual in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual. The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of the condition being treated. Prescription of treatment, eg decisions on dosage etc, is within the responsibility of general practitioners and other medical doctors.

Generally, attenuated *Salmonella* strains and other enteric bacteria (eg *Shigella* and *E.coli*) according to the present invention are provided in an isolated and/or purified form, ie substantially pure. This may include being in a composition where it represents at least about 90% active ingredient, more preferably at least about 95%, more preferably at least about 98%. As stated above, however a composition may include inert carrier materials or other pharmaceutically and physiologically acceptable excipients.

The medicament may be formulated for administration by injection (cutaneous, subcutaneous or intravenous) or for mucosal administration eg via the oral, rectal, nasal, or genital routes.

The *Salmonella* genome or the genome of other enteric bacteria (eg *Shigella* and *E.coli*) as discussed above may itself provide the immunogen or it may contain a heterologous gene insert expressing an immunogenic protein. Thus the present invention can be employed to provide attenuated strains of *Salmonella* and other

enteric bacteria (eg *Shigella* and *E.coli*) for use as safe multivalent vaccines against a variety of foreign pathogens or other antigens for use in man and domestic animals. The attenuated strains can also be employed to provide vaccines against medical conditions such as cancer, by inserting a nucleotide sequence encoding a tumour antigen, for instance, into the genome of an enteric bacterium.

The attenuated *Salmonella* may be an altered form of any serotype of *Salmonella* enterica subspecies enterica. By way of illustration (not limitation) the attenuated *Salmonella* may be an altered strain of *Salmonella typhimurium, Salmonella enteritidis, Salmonella dublin* or *Salmonella choleraesuis*. Many international culture collections hold suitable strains of *Salmonella* which may be altered as described herein to achieve attenuation. In particular *Salmonella* strains are available from the National Collection of Type Cultures, Central Public Health Lab, Collindale Av, Collindale, London NW9. For example: *S.typhimurium* NCTC 12023 (ATCC 14028), *S.enteritidis* NCTC12694 (ATCC13076), *S.dublin* NCTC 9676, *S.choleraesuis* NCTC5735 (ATCC 13312). *Salmonella dublin* 2229 as mentioned later herein is available from the Institute for Animal Health, Compton, Newbury, Berks, United Kingdom RG20 7NN.

The present inventors also provide a plurality of previously uncharacterised useful gene sequences. In particular they provide *sopA, pipD, orfX, pipC, pipB, pipA sopD* and *sipA*.

Thus the present invention also relates to nucleic acid,

polypeptides, oligonucleotide probes and primers, diagnostic, prognostic and other materials and methods relating to these genes. The sequences for *sopA*, *pipD*, *orfX*, *pipC*, *pipB* and *pipA* are all newly provided by the present applicants (see figures 1 to 3 and 6).

The term 'gene' or 'allele' includes normal alleles of the genes disclosed herein (eg *sopA*, *pipD* etc) and also alleles which although carrying one or more variations retain the function of the gene and polypeptide sequences disclosed herein. The term 'nucleic acid' includes a nucleic acid molecule which has a nucleotide sequence encoding a polypeptide which includes an amino acid sequence of a polypeptide as shown in figure 3/figure 6. The coding sequence may be as shown in figure 3/figure 6 or it may be a mutant, variant, derivative or allele of a sequence as shown. The sequence may differ from that shown by a change which is one or more of addition, insertion, deletion and substitution of one or more nucleotides of a sequence shown. Changes to a nucleotide sequence may result in an amino acid change at the protein level, or not, as determined by the genetic code.

Thus, nucleic acid according to the present invention may include a sequence which although different from a sequence shown in figure 3/figure 6 nevertheless encodes a polypeptide with the same amino acid sequence as shown in figure 3/figure 6.

On the other hand, the encoded polypeptide may comprise an amino acid sequence which differs by one or more amino acid residues from an amino acid polypeptide sequence as shown in figure 3/figure 6. Nucleic acid encoding a

polypeptide which is an amino acid sequence mutant, variant, derivative or allele of a polypeptide sequence as shown in figure 3/figure 6 is further provided by the present invention. Such polypeptides are discussed below. Nucleic acid encoding such a polypeptide may show greater than about 20% homology with a coding sequence as shown in figure 3/figure 6, greater than about 30% homology, greater than about 40% homology, greater than about 50% homology, greater than about 60% homology, greater than about 70% homology, greater than about 80% homology, greater than about 90% or greater than about 95% homology. Homologous sequences of requisite identity may be obtained as explained above.

Thus the present invention also provides nucleic acid molecules representing part or all of the *sopA, pipD, pipC, orfX, pipB* or *pipA* gene as set out in figure 3/figure 6 or alleles thereof. Also provided are nucleic acid molecules which have a nucleotide sequence encoding a *sopA, pipD, pipC, orfX, pipB or pipA* amino acid sequence set out in figure 3/figure 6. Provided are nucleic acid molecules which have a nucleotide sequence encoding a polypeptide which is a mutant, variant, derivative or allele of a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide including the amino acid sequence set out in figure 3/figure 6. Provided is nucleic acid encoding a polypeptide having a 20% or more, 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 90% or more or 95% or more sequence homology to *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide amino acid sequence set out in figure 3/figure 6.

Provided are nucleic acid molecules which have a

nucleotide sequence encoding a fragment or active portion of a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide amino acid sequence set out in figure 3/figure 6.

A nucleic acid molecule as provided may comprise a nucleotide sequence functioning as a promoter region.

Generally, nucleic acid according to the present invention maybe provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally associated, such as free or substantially free of nucleic acid normally flanking the gene except possibly one or more regulatory sequence(s) for expression. Nucleic acid may be wholly or partially synthetic and may include genomic DNA, cDNA or RNA. Where nucleic acid according to the invention includes RNA, reference to the sequence shown should be construed as reference to the RNA equivalent, with U substituted for T.

Nucleic acid molecules as provided may comprise a label. A nucleic acid molecule as provided may be for use in a method of medical treatment or diagnosis.

Nucleic acid sequences as disclosed herein can be readily prepared by the skilled person using the information and references contained herein and techniques known in the art (for example, see Sambrook, Fritsch and Maniatis, "Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, and Ausubel et al, Short Protocols in Molecular Biology, John Wiley and Sons, 1992). These techniques include (i) the use of the polymerase chain reaction (PCR) to amplify samples of

such nucleic acid, e.g. from suitable samples comprising *Salmonella, Shigella or E.coli*, (ii) chemical synthesis, or (iii) preparing cDNA sequences. Modifications to the sequences can be made, e.g. using site directed mutagenesis, to lead to the expression of modified polypeptide or to take account of codon preference in the host cells used to express the nucleic acid.

In order to obtain expression of the nucleic acid sequences as provided, the sequences can be incorporated in a vector having control sequences operably linked to control its expression. The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid, nucleic acid sequences so that the polypeptide is produced as a fusion and/or nucleic acid encoding secretion signals so that the polypeptide produced in the host cell is secreted from the cell. Polypeptide can then be obtained by transforming the vectors into host cells in which the vector is functional, culturing the host cells so that the polypeptide is produced and recovering the polypeptide from the host cells or the surrounding medium. Prokaryotic and eukaryotic cells are used for this purpose in the art, including strains of E. coli, yeast, and eukaryotic cells such as COS or CHO cells. The choice of host cell can be used to control the properties of the polypeptide expressed in those cells, e.g. controlling where the polypeptide is deposited in the host cells or affecting properties such as its glycosylation. Thus provided are replicable vectors comprising a nucleic acid as hereby provided operably linked to control sequences to direct its expression; host cells transformed with such a vector; methods of

producing a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide comprising culturing such a host cell so that polypeptide is produced - the method may comprise the further step of recovering the polypeptide produced.

PCR techniques for the amplification of nucleic acid are described in US Patent No. 4,683,195. In general, such techniques require that sequence information from the ends of the target sequence is known to allow suitable forward and reverse oligonucleotide primers to be designed to be identical or similar to the polynucleotide sequence that is the target for the amplification. PCR comprises steps of denaturation of template nucleic acid (if double-stranded), annealing of primer to target, and polymerisation. The nucleic acid probed or used as template in the amplification reaction may be genomic DNA, cDNA or RNA. PCR can be used to amplify specific sequences from genomic DNA, specific RNA sequences and cDNA transcribed from mRNA, bacteriophage or plasmid sequences. The nucleic acid sequences provided herein see for example figure 3 and figure 6 readily allow the skilled person to design PCR primers. References for the general use of PCR techniques include Mullis et al, Cold Spring Harbor Symp. Quant. Biol., 51:263, (1987), Ehrlich (ed), PCR technology, Stockton Press, NY, 1989, Ehrlich et al, Science, 252:1643-1650, (1991), "PCR protocols; A Guide to Methods and Applications", Eds. Innis et al, Academic Press, New York, (1990).

Also included within the scope of the invention are antisense oligonucleotide sequences based on the nucleic acid sequences described herein. Antisense oligonucleotides may be designed to hybridise to the

complementary sequence of nucleic acid, pre-mRNA or mature mRNA, interfering with the production of polypeptide encoded by a given DNA sequence (e.g. either native polypeptide or a mutant form thereof), so that its expression is reduced or prevented altogether. The construction of antisense sequences and their use is described in Peyman and Ulman, Chemical Reviews, 90:543-584, (1990), Crooke, Ann. Rev. Pharmacol. Toxicol., 32:329-376, (1992), and Zamecnik and Stephenson, P.N.A.S, 75:280-284, (1974).

The nucleic acid sequences provided in figure 3 and figure 6 are useful for identifying nucleic acid of interest (and which may be according to the present invention) in a test sample. The present invention provides a method of obtaining nucleic acid of interest, the method including hybridisation of a probe having a sequence shown in figure 3 or figure 6 or a complementary sequence thereto to target nucleic acid.

Hybridisation is generally followed by identification of successful hybridisation and isolation of nucleic acid which has hybridised to the probe, which may involve one or more steps of PCR.

Thus nucleic acid according to the present invention is obtainable using one or more oligonucleotide probes or primers designed to hybridise with one or more fragments of the nucleic acid sequence shown in figure 3 or figure 6, particularly fragments of relatively rare sequence, based on codon usage or statistical analysis. A primer designed to hybridise with a fragment of the nucleic acid sequence shown, may be used in conjunction with one or

more oligonucleotides designed to hybridise to a sequence in a cloning vector within which target nucleic acid has been cloned, or in so-called "RACE" (rapid amplification of cDNA ends) in which cDNA's in a library are ligated to an oligonucleotide linker and PCR is performed using a primer which hybridises with the sequence shown in figure 3 or figure 6 and a primer which hybridises to the oligonucleotide linker.

The conditions of the hybridisation can be controlled to minimise non-specific binding, and preferably stringent to moderately stringent hybridisation conditions are preferred. The skilled person is readily able to design such probes, label them and devise suitable conditions for the hybridisation reactions, assisted by textbooks such as Sambrook et al (1989) and Ausubel et al (1992).

Binding of a probe to target nucleic acid (e.g. DNA) may be measured using any of a variety of techniques at the disposal of those skilled in the art. For instance, probes may be radioactively, fluorescently or enzymatically labelled. Other methods not employing labelling of probe include examination of restriction fragment length polymorphisms, amplification using PCR, RNAase cleavage and allele specific oligonucleotide probing.

Probing may employ the standard Southern blotting technique. For instance nucleic acid material may be extracted from cells and digested with different restriction enzymes. Restriction fragments may then be separated by electrophoresis on an agarose gel, before denaturation and transfer to a nitrocellulose filter.

Labelled probe may be hybridised to the DNA fragments on the filter and binding determined.

Preliminary experiments may be performed by hybridising under low stringency conditions various probes to Southern blots of DNA digested with restriction enzymes. Suitable conditions would be achieved when a large number of hybridising fragments were obtained while the background hybridisation was low.

Those skilled in the art are well able to employ suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on.

On the basis of amino acid sequence information, oligonucleotide probes or primers may be designed, taking into account the degeneracy of the genetic code, and, where appropriate, codon usage of the organism from the candidate nucleic acid is derived. An oligonucleotide for use in nucleic acid amplification may have about 10 or fewer codons (e.g. 6, 7 or 8), i.e. be about 30 or fewer nucleotides in length (e.g. 18, 21 or 24). Generally specific primers are upwards of 14 nucleotides in length, but not more than 18-20. Those skilled in the art are well versed in the design of primers for use processes such as PCR.

A further aspect of the present invention provides an oligonucleotide or polynucleotide fragment of a nucleotide sequence shown in figure 3 or figure 6 or a complementary sequence, in particular for use in a method

of obtaining and/or screening nucleic acid. The sequences referred to above may be modified by addition, substitution, insertion or deletion of one or more nucleotides, but preferably without abolition of ability to hybridise selectively with nucleic acid with a sequence shown in figure 3 or figure 6, that is wherein the degree of homology of the oligonucleotide or polynucleotide with one of the sequences given is sufficiently high.

Thus also provided are methods of identifying a target nucleic acid molecule in a test sample using a nucleic acid probe having all or a portion of a sequence shown in figure 3 or figure 6 or a complementary sequence thereof, the method comprising contacting the probe and the test sample under hybridising conditions and observing whether hybridisation takes place. The probe may be used to identify a *sopA, pipD, pipC, orfX, pipB or pipA* nucleic acid sequence or a mutant allele thereof.

Such probes may be used to identify the presence in the test sample of *Salmonella*, and other enteric bacteria (eg *Shigella* and *E.coli*).

Oligonucleotides according to the present invention that are fragments of any of the sequences shown in figure 3 or figure 6 are at least about 10 nucleotides in length, more preferably at least about 15 nucleotides in length, more preferably at least about 20 nucleotides in length. Such fragments themselves individually represent aspects of the present invention. Fragments and other oligonucleotides may be used as primers or probes as discussed but may also be generated (e.g. by PCR) in

methods concerned with determining the presence in a test sample of a sequence indicative of *Salmonella*, and other enteric bacteria (eg *Shigella* and *E.coli*).

Methods involving use of nucleic acid in diagnostic and/or prognostic contexts, for instance in determining infection with *Salmonella*, and other enteric bacteria (eg *Shigella* and *E.coli*) and other methods concerned with determining the presence of sequences indicative of such infection are discussed below.

Nucleic acid according to the present invention, such as a full-length coding sequence or oligonucleotide probe or primer, may be provided as part of a kit, e.g. in a suitable container such as a vial in which the contents are protected from the external environment. The kit may include instructions for use of the nucleic acid, e.g. in PCR and/or a method for determining the presence of nucleic acid of interest in a test sample. A kit wherein the nucleic acid is intended for use in PCR may include one or more other reagents required for the reaction, such as polymerase, nucleosides, buffer solution etc. The nucleic acid may be labelled. A kit for use in determining the presence or absence of nucleic acid of interest may include one or more articles and/or reagents for performance of the method, such as means for providing the test sample itself, e.g. a swab or a syringe (such components generally being sterile).

A convenient way of producing a polypeptide as disclosed herein is to express nucleic acid encoding it, by use of the nucleic acid in an expression system. The use of expression system has reached an advanced degree of

sophistication today.

Accordingly, the present invention also encompasses a method of making a polypeptide (as disclosed; see below), the method including expression from nucleic acid (generally nucleic acid according to the invention) encoding the polypeptide. This may conveniently be achieved by growing a host cell in culture, containing such a vector, under appropriate conditions which cause or allow expression of the polypeptide. Polypeptides may also be expressed in in vitro systems, such as reticulocyte lysate.

Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, eukaryotic cells such as mammalian and yeast, and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells, COS cells and many others. A common, preferred bacterial host is E. coli.

Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids, viral e.g. 'phage, or phagemid, as appropriate. For further details see, for example, Molecular Cloning: a Laboratory Manual: 2nd edition, Sambrook et al., 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example

in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in Current Protocols in Molecular Biology, Ausubel et al. eds., John Wiley & Sons, 1992.

Thus the present invention provides vectors and host cells containing nucleic acid as disclosed herein. The nucleic acid of the invention may be integrated into the genome (e.g. chromosome) of a host cell. Integration may be promoted by inclusion of sequences which promote recombination with the genome, in accordance with standard techniques. The nucleic acid may be on an extra-chromosomal vector within the cell.

A still further aspect provides a method which includes introducing the nucleic acid into a host cell. The introduction, which may (particularly for *in vitro* introduction) be generally referred to without limitation as "transformation", may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage. As an alternative, direct injection of the nucleic acid could be employed.

Marker genes such as antibiotic resistance or sensitivity genes may be used in identifying clones containing nucleic acid of interest, as is well known in the art.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells (which may include cells actually transformed although more likely the cells will be descendants of the transformed cells) under conditions for expression of the gene, so that the encoded polypeptide is produced. If the polypeptide is expressed coupled to an appropriate signal leader peptide it may be secreted from the cell into the culture medium. Following production by expression, a polypeptide may be isolated and/or purified from the host cell and/or culture medium, as the case may be, and subsequently used as desired, e.g. in the formulation of a composition which may include one or more additional components, such as a pharmaceutical composition which includes one or more pharmaceutically acceptable excipients, vehicles or carriers (e.g. see below).

A host cell containing nucleic acid according to the present invention, e.g. as a result of introduction of the nucleic acid into the cell or into an ancestor of the cell and/or genetic alteration of the sequence endogenous to the cell or ancestor (which introduction or alteration may take place in vivo or ex vivo), may be comprised (e.g. in the soma) within an organism which is an animal.

Instead of or as well as being used for the production of a polypeptide encoded by a transgene, host cells may be used as a nucleic acid factory to replicate the nucleic acid of interest in order to generate large amounts of it. Multiple copies of nucleic acid of interest may be made within a cell when coupled to an amplifiable gene such as DHFR. Host cells transformed with nucleic acid

of interest, or which are descended from host cells into which nucleic acid was introduced, may be cultured under suitable conditions, e.g. in a fermenter, taken from the culture and subjected to processing to purify the nucleic acid. Following purification, the nucleic acid or one or more fragments thereof may be used as desired, for instance in a diagnostic or prognostic assay as discussed elsewhere herein.

The skilled person can use the techniques described herein and others well known in the art to produce large amounts of the *sopA, pipD, orfX, pipC, pipB and pipA* polypeptide, or fragments or active portions thereof, for use as pharmaceuticals, in the developments of vaccines and for further study into its properties and role in vivo.

A further aspect of the present invention provides a polypeptide which has an amino acid sequence shown in figure 3 or figure 6, which may be in isolated and/or purified form, free or substantially free of material with which it is naturally associated.

Thus the present invention also provides a substance which is a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide encoded by a nucleic acid molecule as stated above. Also provided is a substance which is a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide including the amino acid sequence set out in figure 3 or figure 6. Also provided is a substance which is a polypeptide having 20% or more, 30% or more, 40% or more, 50% or more, or 60% or more, 70% or more, 80% or more, 90% or more or 95% or more sequence homology to a *sopA, pipD,*

pipC, orfX, pipB or pipA polypeptide including an amino acid sequence set out in figure 3 or figure 6. Also provided is a substance which is a polypeptide mutant, variant, derivative or allele of a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide as set out in figure 3 or figure 6. A polypeptide which is a variant, allele, derivative or mutant may have an amino acid sequence which differs from that given in figure 3 or figure 6 by one or more of addition, substitution, deletion and insertion of one or more amino acids. Preferred such polypeptides have a function broadly equivalent to the native polypeptides, that is to say have one or more of the following properties: immunological cross-reactivity with an antibody reactive with a polypeptide for which the sequence is given in figure 3 or figure 6; sharing an epitope with a polypeptide for which the amino acid sequence is shown in figure 3 or figure 6 (as determined for example by immunological cross-reactivity between the two polypeptides).

A polypeptide which is an amino acid sequence variant, allele, derivative or mutant of an amino acid sequence shown in figure 3 or figure 6 may comprise an amino acid sequence which shares greater than about 20% sequence identity, greater than about 30%, greater than about 40%, greater than about 50%, greater than about 60%, greater than about 70%, greater than about 80%, greater than about 90% or greater than about 95%. The sequence may share greater than about 50% similarity, greater than about 60% greater than about 70% similarity, greater than about 80% similarity or greater than about 90% similarity with an amino acid sequence shown in figure 3 or figure 6. Particular amino acid sequence variants may differ

from those shown in figure 3 or figure 6 by insertion, addition, substitution or deletion of 1 amino acid, 2, 3, 4, 5-10, 10-20 20-30, 30-50, 50-100, 100-150, or more than 150 amino acids.

The present invention also includes active portions, fragments, derivatives and functional mimetics of the polypeptides provided herein.

An "active portion" of a *sopA, pipD, orfX, pipC, pipB and pipA* polypeptide means a peptide which is less than the full length polypeptide, but which retains its essential biological activity. A "fragment" of the polypeptide means a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. Fragments may comprise the antigenic determinants or epitopes characteristic of the wild-type full length polypeptide which are useful for raising antibodies.

A "derivative" of the polypeptide or a fragment thereof means a polypeptide modified by varying the amino acid sequence of the protein, e.g. by manipulation of the nucleic acid encoding the protein or by altering the protein itself. Such derivatives of the natural amino acid sequence may involve insertion, addition, deletion or substitution of one or more amino acids, without fundamentally altering the essential activity of the wild type polypeptide.

Thus provided are substances which are fragments or active portions or functional mimetics of a *sopA*, *pipD*, *pipC*, *orfX*, *pipB* or *pipA* polypeptide including an amino acid sequence of figure 3 or figure 6.

"Functional mimetic" means a substance which may not contain an active portion of the native amino acid sequence, and probably is not a peptide at all, but which retains the essential biological activity of natural polypeptide. The design and screening of candidate mimetics is described in detail below.

A polypeptide according to the present invention may be isolated and/or purified (e.g. using an antibody) for instance after production by expression from encoding nucleic acid. Polypeptides according to the present invention may also be generated wholly or partly by chemical synthesis using the sequence information provided. The isolated and/or purified polypeptide may be used in formulation of a composition, which may include at least one additional component, for example a vaccine composition including a pharmaceutically acceptable excipient, vehicle or carrier. A composition including a polypeptide according to the invention may be used in prophylactic and/or therapeutic treatment as discussed herein.

A polypeptide, peptide fragment, allele, mutant or variant according to the present invention may be used as an immunogen or otherwise in obtaining antibodies that have the property of specifically binding to the polypeptides, fragments, alleles, mutants, variants or active portions thereof. Antibodies are useful in

purification and other manipulation of polypeptides and peptides, diagnostic screening and therapeutic contexts. This is discussed further herein.

A polypeptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic context.

The production of monoclonal antibodies is well established in the art. Monoclonal antibodies can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP-A-184187, GB-A-2188638 or EP-A-239400. A hybridoma producing a monoclonal antibody may be subject to genetic mutation or other changes, which may or may not alter the binding specificity of antibodies produced.

The provision of the novel polypeptides enables for the first time the production of antibodies able to bind specifically to them. Accordingly, a further aspect of the present invention provides an antibody able to bind specifically to a polypeptide as hereby provided eg a polypeptide whose sequence is given in figure 3. Such an antibody may be specific in the sense of being able to distinguish between the polypeptide it is able to bind and other polypeptides for which it has no or

substantially no binding affinity (e.g. a binding affinity of about 1000x worse). Specific antibodies bind an epitope on the molecule which is either not present or is not accessible on other molecules. Antibodies according to the present invention may be specific for the wild-type polypeptide. Antibodies according to the invention may be specific for a particular mutant, variant, allele or derivative polypeptide as between that molecule and the wild-type polypeptide, so as to be useful in diagnostic and prognostic methods. Antibodies are also useful in purifying the polypeptide or polypeptides to which they bind, e.g. following production by recombinant expression from encoding nucleic acid.

Thus provided are antibodies capable of specifically binding to a *sopA, pipD, pipC, orfX, pipB or pipA* polypeptide as above. Antibodies may be labelled. They may be used to determine the presence, amount or location in a sample of the polypeptides above or *Salmonella*, and other enteric bacteria (eg *Shigella* and *E.coli*).

Preferred antibodies according to the invention are isolated, in the sense of being free from contaminants such as antibodies able to bind other non-related polypeptides and/or free of serum components. Monoclonal antibodies are preferred for some purposes, though polyclonal antibodies are within the scope of the present invention.

Antibodies may be obtained using techniques which are standard in the art. Methods of producing antibodies include immunising a mammal (e.g. mouse, rat, rabbit,

horse, goat, sheep or monkey) with the protein or a fragment thereof. Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al, Nature, 357:80-82, 1992). Isolation of antibodies and/or antibody-producing cells from an animal may be accompanied by a step of sacrificing the animal.

As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a protein may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047. The library may be naive, that is constructed from sequences obtained from an organism which has not been immunised with any of the proteins (or fragments), or may be one constructed using sequences obtained from an organism which has been exposed to the antigen of interest.

Antibodies according to the present invention may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any binding substance having a binding domain with the required specificity. Thus the invention covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including synthetic molecules and molecules whose shape mimics that of an antibody enabling it to bind an antigen or epitope.

Example antibody fragments, capable of binding an antigen or other binding partner are the Fab fragment consisting of the VL, VH, Cl and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an antibody; the dAb fragment which consists of a VH domain; isolated CDR regions and F(ab')2 fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

Humanised antibodies in which CDRs from a non-human source are grafted onto human framework regions, typically with the alteration of some of the framework amino acid residues, to provide antibodies which are less immunogenic than the parent non-human antibodies, are also included within the present invention

A hybridoma producing a monoclonal antibody according to the present invention may be subject to genetic mutation or other changes. It will further be understood by those skilled in the art that a monoclonal antibody can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP-A-184187, GB-A-2188638 or EP-A-0239400. Cloning and expression of chimeric antibodies are described in EP-A-0120694 and EP-A-0125023.

Hybridomas capable of producing antibody with desired binding characteristics are within the scope of the present invention, as are host cells, eukaryotic or prokaryotic, containing nucleic acid encoding antibodies (including antibody fragments) and capable of their expression. The invention also provides methods of production of the antibodies including growing a cell capable of producing the antibody under conditions in which the antibody is produced, and preferably secreted.

The reactivities of antibodies on a sample may be determined by any appropriate means. Tagging with individual reporter molecules is one possibility. The reporter molecules may directly or indirectly generate detectable, and preferably measurable, signals. The linkage of reporter molecules may be directly or indirectly, covalently, e.g. via a peptide bond or non-covalently. Linkage via a peptide bond may be as a result of recombinant expression of a gene fusion encoding antibody and reporter molecule.

One favoured mode is by covalent linkage of each antibody with an individual fluorochrome, phosphor or laser dye with spectrally isolated absorption or emission characteristics. Suitable fluorochromes include fluorescein, rhodamine, phycoerythrin and Texas Red. Suitable chromogenic dyes include diaminobenzidine.

Other reporters include macromolecular colloidal particles or particulate material such as latex beads that are coloured, magnetic or paramagnetic, and biologically or chemically active agents that can directly or indirectly cause detectable signals to be

visually observed, electronically detected or otherwise recorded. These molecules may be enzymes which catalyse reactions that develop or change colours or cause changes in electrical properties, for example. They may be molecularly excitable, such that electronic transitions between energy states result in characteristic spectral absorptions or emissions. They may include chemical entities used in conjunction with biosensors. Biotin/avidin or biotin/streptavidin and alkaline phosphatase detection systems may be employed.

The mode of determining binding is not a feature of the present invention and those skilled in the art are able to choose a suitable mode according to their preference and general knowledge.

Antibodies according to the present invention may be used in screening for the presence of a polypeptide according to the present invention, for example in a test sample containing cells or cell lysate as discussed, and may be used in purifying and/or isolating a polypeptide according to the present invention, for instance following production of the polypeptide by expression from encoding nucleic acid therefor. Antibodies may modulate the activity of the polypeptide to which they bind and so, if that polypeptide has a deleterious effect in an individual, may be useful in a therapeutic context (which may include prophylaxis).

An antibody may be provided in a kit, which may include instructions for use of the antibody, e.g. in determining the presence of a particular substance in a test sample. One or more other reagents may be included, such as

labelling molecules, buffer solutions, elutants and so on. Reagents may be provided within containers which protect them from the external environment, such as a sealed vial.

A number of methods are known in the art for analysing biological samples from individuals to determine whether the individual carry an undesirable bacterial species such as *Salmonella*, and other enteric bacteria (eg *Shigella* and *E.coli*).

Broadly, the methods divide into those screening for the presence of nucleic acid sequences characteristic of the virus and those that rely on detecting the presence or absence of polypeptide or characteristic antibody thereto. The methods make use of biological samples from individuals that are suspected of contain the nucleic acid sequences or polypeptide. Examples of biological samples include blood, plasma, serum, tissue samples.

In most screening methods based on analysis of nucleic acid, nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other sequences present in the sample. This allows the target sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

There are various methods for determining the presence or absence in a test sample of a particular polypeptide, such as a polypeptide with an amino acid sequence shown

in figure 3 or figure 6, or an amino acid sequence mutant, variant or allele thereof.

A sample may be tested for the presence of a binding partner for a specific binding member such as an antibody (or mixture of antibodies), specific for one or more particular variants of a polypeptide shown in figure 3 or figure 6.

A sample may be tested for the presence of a binding partner for a specific binding member such as an antibody (or mixture of antibodies), specific for a polypeptide shown in figure 3 or figure 6.

In such cases, the sample may be tested by being contacted with a specific binding member such as an antibody under appropriate conditions for specific binding, before binding is determined, for instance using a reporter system. Where a panel of antibodies is used, different reporting labels may be employed for each antibody so that binding of each can be determined.

A specific binding member such as an antibody may be used to isolate and/or purify its binding partner polypeptide from a test sample, to allow for sequence and/or biochemical analysis of the polypeptide to determine whether it has the sequence and/or properties of a polypeptide whose sequence is shown in figure 3 or figure 6, or if it is a mutant or variant form thereof. Amino acid sequence is routine in the art using automated sequencing machines.

There is also an increasing tendency in the diagnostic

field towards miniaturisation of such assays, e.g. making use of binding agents (such as antibodies or nucleic acid sequences) immobilised in small, discrete locations (microspots) and/or as arrays on solid supports or on diagnostic chips. These approaches can be particularly valuable as they can provide great sensitivity (particularly through the use of fluorescent labelled reagents), require only very small amounts of biological sample from individuals being tested and allow a variety of separate assays can be carried out simultaneously. Examples of techniques enabling this miniaturised technology are provided in WO84/01031, WO88/1058, WO89/01157, WO93/8472, WO95/18376/ WO95/18377, WO95/24649 and EP-A-0373203. Thus, in a further aspect, the present invention provides a kit comprising a support or diagnostic chip having immobilised thereon one or more binding agents capable of specifically binding nucleic acid as provided herein optionally in combination with other reagents (such as labelled developing reagents) needed to carrying out an assay.

The mutant bacteria, polypeptides, antibodies, peptides and nucleic acid of the invention can be formulated in pharmaceutical compositions for the prevention or treatment of infection by *Salmonella,* and other enteric bacteria (eg *Shigella* and *E.coli*). For vaccine use the gene products (wild type or genetically modified/detoxified) can be overexpressed in an attenuated carrier strain. Thus also provided are pharmaceutical compositions. These compositions may comprise, in addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabiliser or other materials well known to those skilled

in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

Pharmaceutical compositions for oral administration may be in tablet, capsule, powder or liquid form. A tablet may include a solid carrier such as gelatin or an adjuvant. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included.

For intravenous, cutaneous or subcutaneous injection, the active ingredient will be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection, Lactated Ringer's Injection. Preservatives, stabilisers, buffers, antioxidants and/or other additives may be included, as required.

Whether it is a mutant bacterium, polypeptide, antibody, peptide, nucleic acid molecule, small molecule or other pharmaceutically useful compound according to the present invention that is to be given to an individual, administration is preferably in a "prophylactically

effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual. The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of what is being treated. Prescription of treatment, e.g. decisions on dosage etc, is within the responsibility of general practitioners and other medical doctors, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of delivery, the method of administration and other factors known to practitioners. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 16th edition, Osol, A. (ed), 1980.

A polypeptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic (possibly including prophylactic) context.

It is well known that pharmaceutical research leading to the identification of a new drug may involve the screening of very large numbers of candidate substances, both before and even after a lead compound has been found. This is one factor which makes pharmaceutical research very expensive and time-consuming. Means for assisting in the screening process can have considerable commercial importance and utility. Such means for screening for substances potentially useful in treating infection by *Salmonella*, and other enteric bacteria (eg

Shigella and *E.coli*) is provided by polypeptides according to the present invention. Substances identified as modulators of the polypeptide will provide basis for design and investigation of therapeutics for in vivo use.

A method of screening for a substance which modulates activity of a polypeptide may include contacting one or more test substances with the polypeptide in a suitable reaction medium, testing the activity of the treated polypeptide and comparing that activity with the activity of the polypeptide in comparable reaction medium untreated with the test substance or substances. A difference in activity between the treated and untreated polypeptides is indicative of a modulating effect of the relevant test substance or substances.

Combinatorial library technology provides an efficient way of testing a potentially vast number of different substances for ability to modulate activity of a polypeptide. Such libraries and their use are known in the art. The use of peptide libraries is preferred.

Prior to or as well as being screened for modulation of activity, test substances may be screened for ability to interact with the polypeptide, e.g. in a yeast two-hybrid system (which requires that both the polypeptide and the test substance can be expressed in yeast from encoding nucleic acid). This may be used as a coarse screen prior to testing a substance for actual ability to modulate activity of the polypeptide.

Following identification of a substance which modulates

or affects polypeptide activity, the substance may be investigated further. Furthermore, it may be manufactured and/or used in preparation, i.e. manufacture or formulation, of a composition such as a medicament, pharmaceutical composition or drug. These may be administered to individuals.

Thus, the present invention extends in various aspects not only to a substance identified using a nucleic acid molecule as a modulator of polypeptide activity, in accordance with what is disclosed herein, but also a pharmaceutical composition, medicament, drug or other composition comprising such a substance, a method comprising administration of such a composition to a patient, e.g. for treatment (which may include preventative treatment) of infection by *Salmonella*, and other enteric bacteria (eg *Shigella* and *E.coli*), use of such a substance in manufacture of a composition for administration and a method of making a pharmaceutical composition comprising admixing such a substance with a pharmaceutically acceptable excipient, vehicle or carrier, and optionally other ingredients.

A substance identified using as a modulator of polypeptide function may be peptide or non-peptide in nature. Non-peptide "small molecules" are often preferred for many *in vivo* pharmaceutical uses. Accordingly, a mimetic or mimic of the substance (particularly if a peptide) may be designed for pharmaceutical use.

The designing of mimetics to a known pharmaceutically active compound is a known approach to the development of

pharmaceuticals based on a "lead" compound. This might be desirable where the active compound is difficult or expensive to synthesise or where it is unsuitable for a particular method of administration, e.g. peptides are unsuitable active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing is generally used to avoid randomly screening large number of molecules for a target property.

There are several steps commonly taken in the design of a mimetic from a compound having a given target property. Firstly, the particular parts of the compound that are critical and/or important in determining the target property are determined. In the case of a peptide, this can be done by systematically varying the amino acid residues in the peptide, eg by substituting each residue in turn. Alanine scans of peptide are commonly used to refine such peptide motifs. These parts or residues constituting the active region of the compound are known as its "pharmacophore".

Once the pharmacophore has been found, its structure is modelled to according its physical properties, eg stereochemistry, bonding, size and/or charge, using data from a range of sources, eg spectroscopic techniques, X-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modelling process.

In a variant of this approach, the three-dimensional

structure of the ligand and its binding partner are modelled. This can be especially useful where the ligand and/or binding partner change conformation on binding, allowing the model to take account of this in the design of the mimetic.

A template molecule is then selected onto which chemical groups which mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted on to it can conveniently be selected so that the mimetic is easy to synthesise, is likely to be pharmacologically acceptable, and does not degrade *in vivo*, while retaining the biological activity of the lead compound. Alternatively, where the mimetic is peptide based, further stability can be achieved by cyclising the peptide, increasing its rigidity. The mimetic or mimetics found by this approach can then be screened to see whether they have the target property, or to what extent they exhibit it. Further optimisation or modification can then be carried out to arrive at one or more final mimetics for *in vivo* or clinical testing.

Thus nucleotides and polypeptides as provided herein have several clear applications *inter alia*:

a. PCR primers can be designed for the detection of *Salmonellas* or *Salmonellas* and *Shigellas*.

b. The gene products themselves may be used as sub-unit vaccines for the control of *Salmonellosis*. Polyclonal antibodies against eg *sopD* have been obtained indicating the immunogenecity of the *sopD* polypeptide.

c. Antibodies eg anti-*sopD* may be utilised for immunodiagnostics or treatment of *Salmonellosis*. The *Sop* proteins were isolated as described earlier (Wood et al., 1996). The proteins were separated by SDS-PAGE and stained with Coomassie blue. *SopD* band was excised from the gel, and the acrylamide was finely dispersed in a Griffin's tube in a minimal volume of sterile PBSa. This suspension was then used to inoculate rabbits. Similarly monoclonal antibodies using the same technique as described for the SopE protein (Wood et al., 1996) have been raised against *sopB* and *sipA*. The ability to raise antibodies indicates the immunogenicity of these gene products. An antibody based immunodiagnostic assay may be developed in accordance with the teachings of Pal et al., 1997, J.Clin.Microbiol, 35 1757-1760 relating to an assay based on the functionally related *Shigella* protein IpaC.

To assist a fuller understanding of the present invention, there now follows a presentation of how the materials and methods of the present invention may be employed and experimental studies related thereto.

Brief description of the drawings

Reference is made to the following figures described below.

Figure 1. Schematic representation of physical and genetic map of SPI-5.

Figure 2. Sequence analysis of junctions between *S.dublin* DNA homologous and non-homologous to *E.coli* K-12 DNA.

Figure 3. Nucleotide sequence of SPI-5 with putative amino acid sequences of encoded proteins.

Figure 4. Effect of different mutations within SPI-5 on the ability of *S.dublin* to induce fluid secretion and PMN influx 12 h after infection of bovine ligated ileal loops. The secretory response is defined as the volume of fluid within a loop per length of loop and the PMN influx is defined as the radioactive count of PMNs within test loops per radioactive count of PMNs within the negative control loops. Each bar is derived from the mean from four loops and is presented with the standard error of the mean. The decreased enteropathogenicity in bovine ileal loops correlates with reduced oral virulence of mutants for calves.

Figure 5. PCR analyses of different bacterial strains using SPI-5-specific primers. Lane 0 contain DNA markers, the other lanes contain PCR products amplified using the following bacterial suspensions as templates: 1. *S.anatum* 997; 2. *S.choleraesuis* A57; 3. *S.dublin* 2229; 4. *S.enteritidis* 149; 5. *S.gallinarum* 289/91; 6. *S.pullorum* 449/87; 7. *S.typhimurium* C5; 8. *Shigella sonnei* PB1; 9. *E.coli* H310 (EPEC); 10. *Yersinia pseudotuberculosis* YPIII/pIB102. The primers used were:
A) PA1 (5'-caggcgaagcctttgaaac-3') and PA2 (5'-ctgggaggtgaaggatgcc-3'), *pipA*-specific.
B) PB1 (5'-cgcaataaagatctcgatctacagacggt-3') and PB2 (5'-tgccgcttttcgaattccaggcgggt-3'), *sopB*-specific.
C) PD1 (5'-cctattatcagggcatgcgtcattatcaac-3') and PD2 (5'-catataatcctcgagagctctgct-3'), *pipD*-specific.

Figure 6. Nucleotide sequence of *sopA* with putative amino

acid sequence of encoded protein.

Figure 7. Effect of mutation in *sopA* on the ability of *S.dublin* to induce fluid secretion and inflammatory response (PMN influx) 12h after infection of bovine ligated ileal loops. The secretory response and PMN influx are as defined above in relation to Figure 4.

Detailed description of the drawings and exemplification of the invention

Experimental procedures

Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table 1. Strains were grown in LB broth or on LB agar with the addition of appropriate antibiotics when necessary.

DNA Methods

Preparation of plasmid DNA, restriction enzyme digestion, ligations, and transformation of *E.coli* were performed essentially as described by Sambrook et al. (1989). pMW60 was used as a template for DNA sequencing using a Sequenase kit, version 2 (USB Corp.) according to the manufacturer's instructions.

Screening of a S.dublin DNA library and cloning of SPI-5 DNA

Screening of a library of chromosomal DNA of *S.dublin* and Southern blots were performed using PCR amplified DNA fragments A to D (Fig.1). The following primers were used in PCR:

Probe A:
PRA1 (5'-ccgatcatacactggaaatgtgc-3'), and
PRA2 (5'-aagtaatatcacccgtgaataa-3').

Probe B:
PRB1 (5'-cggtcttactcgagcggggctttgct-3'), and
PRB2 (5'-gggttaacaacccttaaaagg-3').

Probe C:
PRC1 (5'-taacgcgattgctctattacac-3'), and
PRC2 (5'-tattctttgccgcgatcctttat-3').

Probe D:
PRD1 (5'-ttcctgctcctctggttac-3'), and
PRD2 (5'-ccggacggtcataatgaga-3').

Construction of S.dublin mutants

Mutants of *S.dublin* were constructed as follows.

***S.dublin* PA1** (insertion of a suicide vector into *pipA*): An internal DNA fragment of *pipA* amplified by PCR using *S.dublin* 2229 chromosomal DNA as a template and two oligonucleotide primers, PIPA1 (5'-cacctacagatctatacctcaaagcggagt-3'), and PIPA2 (5'-cagcgctgctcgagatcatgtagttcttt-3'). This DNA fragment was then cloned into the suicide plasmid vector pDM4 (Milton et al., 1996). The resulting plasmid pPA1 was conjugated from *E.coli* S17.1 into *S.dublin* 2229 and Cm^R transconjugants were obtained. The insertion of the suicide vector by a single recombination event into the chromosome of *S.dublin* 2229 resulted in a *pipA* mutant, which was denoted *S.dublin* PA1. *S.dublin* PD1 (insertion of suicide vector into *pipD*) and *S.dublin* KIN (insertion of suicide vector into *copS*) were constructed as

described for *S.dublin* PA1. PIPD1 (5'-ccttcattctcgagcgcaacgaagat-3') and PIPD2 (5'-caggatcacggatccaatggcgtcttccgt-3') primers were used to amplify an internal fragment of *pipD*; COPS1 (5'-cgcagaatggttaagttatctagaacgctg-3') and COPS2 (5'-ggcattcgtcagcaggctcgagagcactctttg-3') primers were used to amplify an internal fragment of *copS*.

***S.dublin* PB2** (in-frame deletion in *pipB*).

Four oligonucleotide primers,
SPIPB1 (5'-taaaatatttctcgaggatgtcaacgggt-3'),
SPIPB2 (5'-AGTCTGAGCCGtttgtttcacggaaatct-3'),
SPIPB3 (5'-CCGTGAAACAAAcggctcagacttaactgac-3') and
SPIPB4 (5'-cacatacagatctcgctatgattcagac-3')
were designed to generate a fusion DNA fragment covering the 5' and 3' area of *pipB* and carrying a 156 bp long deletion of the internal part of the gene. Two DNA fragments were amplified by PCR using *S.dublin* 2229 chromosomal DNA as a template and SPIPB1 and SPIPB2 and SPIPB3 and SPIPB4 respectively. These DNA fragments were then used as a template as a template in a PCR reaction with SPIPB1 and SPIPB4. The resulting DNA fragment was cloned into the suicide plasmid vector pDM4 (Milton et al., 1996) to yield pDPIPB1. pDPIPB1 was conjugated from *E.coli* S17.1 into *S.dublin* 2229 and Cm^r transconjugants were obtained. The suicide plasmid was then excised by a second recombination event as described in (Milton et al., 1996). The Cm sensitive recombinants were obtained and screened by PCR for a mutated allele. Several clones carrying the deletion were identified. One of these was designated *S.dublin* PB2 and used in further experiments.

S.dublin SD1 (insertion of a suicide vector into *sopD*)
Based on the known nucleotide sequence of the ORF4 of *S.typhimurium* (Ostrowski et al., J.Biol.Chem, (1989) 264(26) 15726-37) two oligonucleotide primers SD1 (5'TTA TGC CAG TCG ACT TAA GCT TCG GTA AT-3'), SD2 (5'-TGA TAG TAA ACA GAT CTT GAT GAG C-3') were designed to generate the DNA fragment containing an internal part of the gene. The corresponding DNA fragment was amplified by PCR using *S.dublin* 2229 chromosomal DNA as template. This DNA fragment was used as a probe to screen a plasmid library of *S.dublin* chromosomal DNA. Two positive clones were identified, one of which was denoted pMW45 and chosen for sequencing and transcomplementation experiments.

The *sopD* mutants of *S.dublin* were constructed as follows. An internal DNA fragment of *sopD* was amplified by PCR with SD1 and SD2 using and was cloned into the suicide plasmid vector pDM4. The resulting plasmid was conjugated from *E.coli* S17.1 into *S.dublin* 2229 and Cml^r transconjugants were obtained. One of these clones was denoted *S.dublin* SD1 and chosen for further experiments. The correct insertion of the suicide vector into the *sopD* gene was confirmed by the PCR.

S.dublin SB2SD1
The *sopB/sopD* mutant was constructed as follows. The *sopB* mutant was constructed as described earlier (Galyov et al., 1997 Molecular Microbiology, 25, 903-12). The *sopD* mutation was then transfected by P22 transduction into the *sopB* mutant strain *S.dublin* SB2 to yield *S.dublin* SB2SD1 mutant strain.

S.dublin A1

SipA mutant was constructed as described previously (Wood et al., 1996 M.Microbiol. 22: 327-338).

The mutant forms of *Salmonella* were (as described above) prepared by use of the suicide plasmid technique. This technique is widely used. Those working in the field would also be able to make mutant forms as disclosed herein by other techniques. The mutations may comprise point mutations, insertions, deletions, substitutions at one or more positions in the nucleotide sequences.

Bovine ligated ileal loop assay for enteropathogenesis
This assay has been described in detail elsewhere (Wallis et al., 1995). The bacterial strains were grown overnight at 25°C, with shaking. The cultures were diluted approximately 1 in 3 in fresh LB medium and incubated at 37°C for 90 min with shaking. The optical density was adjusted by adding LB broth to give approximately 9.0 $\log_{10}$cfu ml^{-1}. Sterile LB broth was used as the negative control.

Ligated ileal loops were constructed in the mid-ileum of 28-day-old, male, Friesian calves. The bacterial culture (1 ml) was injected into loops of 6 cm in length. Approximately 50 ml of blood were removed from the calf and the PMNs were isolated, labelled with ^{111}In and reinjected into the calf. Twelve hours after inoculation of the loops, the secretory response (volume of fluid within a loop/length of a loop) was recorded. The γ-irradiation from each loop and its contents were measured and used to calculate the PMN influx ratio (irradiation from test loop/irradiation from negative control loop).

Results

Identification and analysis of the DNA region adjacent to and downstream of sopB.

The inventors have previously identified a ~3 kb *Sal*I-*Eco*RI DNA fragment which hybridised to the *sopB* probe and cloned it into the pBluescript plasmid vector to yield pMW60 (Galyov et al., 1997). To analyse the DNA upstream and downstream of *sopB*, the entire *Sal*I-*Eco*RI fragment was sequenced. The sequence of the fragment was compared with the DNA sequences deposited in the *E.coli* K-12 genomic database by using the BLAST program. This comparison revealed no extensive similarities, indicating that DNA homologous to *sopB* and the flanking DNA is absent from the *E.coli* K-12 chromosome.

In order to gain more insight into the organisation of the genetic loci adjacent to *sopB*, the inventors used PCR amplified DNA fragments corresponding to the regions upstream and downstream of *sopB* as probes to identify clones from a *S.dublin* chromosomal DNA library containing the extended DNA fragments corresponding to the areas upstream and downstream of *sopB*. This analysis revealed no clones hybridising with the DNA probe derived upstream of *sopB*. One clone hybridising with the probe derived downstream of *sopB* was identified and the plasmid from this clone was designated pMW61 (Fig. 1). The 400 bp long nucleotide sequence of the distal end of the cloned fragment in pMW61 was obtained and compared with those in the *E.coli* K-12 genomic database. This analysis revealed that DNA cloned into pMW61 included a gene identical to the $tRNA_1^{Ser}$ (*serT*) gene from *E.coli* (Fig.2). A strong homology between *S.dublin* sequence and the corresponding

sequence from *E.coli* K-12 extended from the *serT* promoter area (end of the cloned fragment), included the *serT* gene, a short fragment downstream of *serT* and a part of an inverted repeat terminator structure downstream of the gene (Fig.2). This was followed by *Salmonella* specific DNA, suggesting that the enteropathogenicity associated *sopB* gene is located on the *Salmonella*-specific DNA fragment absent from the *E.coli* K-12 chromosome and that the site of the insertion for this fragment is the inverted repeat sequence downstream of *serT*. Thus, this *Salmonella*-specific DNA fragment has features characteristic to "pathogenicity islands" (for a review see Groisman and Ochman, 1996), and has now been denoted SPI-5 (previously denoted SPI-4 by the inventors; see earlier comment).

Cloning and analysis of the DNA regions adjacent to and upstream of the sopB gene.

Since the inventors were unable to identify any clones which hybridised with DNA upstream of *sopB* in the available *S.dublin* DNA library, they performed a Southern hybridisation of different restriction enzyme digests of *S.dublin* chromosomal DNA using a probe derived from flanking DNA upstream of *sopB*. A -2.5 kb *Hpa*I DNA fragment which hybridised to the probe was identified and cloned into the pBluescript plasmid vector to yield pMW59. The 300 bp long nucleotide sequence of the *sopB* distal end of the cloned fragment in pMW59 were obtained and compared with those in the *E.coli* K-12 genomic database. Nucleotide sequencing of the sopB distal end of the fragment inserted into pMW59 revealed sequence strongly homologous to the *copS* gene of *E.coli* K-12 (Fig.2). The *copS* gene is located at 46 centisome of the

E.coli chromosome. Two oligonucleotide primers were designed based on the determined sequence, a corresponding DNA fragment was amplified by PCR and used as a probe to identify another chromosomal DNA fragment suitable for cloning DNA extending further upstream of *sopB*. Using this strategy, the inventors obtained a set of overlapping clones extending approximately 8 kb upstream of *sopB* (Fig.1). Further sequence analysis of the DNA fragment cloned into pMW58 and pMW57 revealed that the *copS* gene is followed by the *copR* and ORF genes (Fig.1). The *copS*, *copR* and ORF genes are structurally similar to the corresponding genes from *E.coli* and the relative gene order and orientation is conserved in *S.dublin* and *E.coli* K-12. Thus, SPI-5 of *S.dublin* appears to be flanked by the *serT* on one side and *copS* followed by *copR* and ORF genes on the other side (Fig.1).

Structure of SPI-5

Nucleotide sequence of the DNA fragment flanked by the *serT* gene on one side and *copS* on the other revealed that, in addition to *sopB*, five other ORFs are located within SPI-5 (Fig.1). The inventors denoted corresponding genes as *pipA*, *pipB*, *pipC*, *pipD* (pathogenicity island encoded proteins) and orfX (Fig.1).

The *pipA* gene product does not appear to have significant homology to previously described proteins, whereas the other SPI-5 encoded gene products showed sequence similarity to gene products from other bacterial species (Fig.1). The inventors have previously shown that the *sopB* gene product is homologous to IpgD protein from *Shigella* (Allaoui et al., 1993). The predicted protein product *pipC* is similar to the IpgE proteins of *Shigella*

(Allaoui et al., 1993). The *sopB* and *pipC* are likely to form an operon, similar in organisation to the *ipgD/ipgE* operon from *Shigella* (Allaoui et al., 1993). *pipB* encodes a putative membrane associated protein with two transmembrane helices. The *pipB* gene product is structurally similar to the Hg1K protein from *Anabaena* and *Synechocystis* species (Black et al., 1995), as well as to several other putative membrane proteins from *Synechocystis*. The *pipD* gene product is a structural homolog to dipeptidases from two different *Lactobacillus* species. The gene product of *orfX* is similar to a putative protein ECU82528 from *E.coli*

The G+C content of SPI-5 is 46%, compared to 44.6% for SPI-2 (Shea et al., 1996), 47% for SPI-1, 39.8-49.3% for SPI-3 (Blanc-Potard and Groisman, 1997), and 52-54% estimated for the entire *Salmonella* genome (ref).

SPI-5 is conserved in Salmonellae

A 1-kb *sopB*-specific DNA fragment was amplified by PCR and used as a probe in Southern blot hybridisation analysis of genomic DNA from different *Salmonella* serovars and some other pathogenic bacteria. DNA fragment hybridisation showed that SPI-5 is present in *S.typhimurium, S.enteritidis, S.cholerae-suis, S.gallinarum, S.pullorum,* and is absent from *Yersinia pseudotuberculosis, Shigella flexneri,* and EPEC (data not shown).

The position of the SPI-5 was determined by screening a prophage library of the *S.typhimurium* genome by PCR using primers to the genes *sopB, pipD, serT* and the junction of *copS/copR*. The PCR products of the expected sizes were

obtained when a template corresponding to 17 to 21.5 centisomes on the chromosome was present (data not shown).

Mutational analysis

The inventors have previously demonstrated that the SopB protein is involved in enteropathogenicity of *Salmonella* (Galyov et al., 1997). To investigate the possible involvement of other SPI-5 encoded genes in the induction of intestinal fluid secretion and inflammatory responses mediated by *Salmonella*, they constructed mutants in the *pipA, pipB, pipD and copS* genes and assessed corresponding mutant strains for growth in vitro, invasiveness of HeLa cells, and enteropathogenicity in bovine ileal ligated loops. In addition, the wild-type *S.dublin* 2229 and a *sopB* mutant *S.dublin* SB2 (Galyov et al., 1997) were also included as controls in the experiments.

The growth characteristics of the SPI-5, *sipA* and *sopD pip* mutant strains in vitro were undistinguishable from that of the wild-type. In contrast, the *copS* mutant showed reduced viability on McCconkey plates (data not shown). None of the mutations affected the ability of *Salmonella* to invade cultured HeLa cells (data not shown). However, the *pipA, pipB, pipD, sopB, sopD* and *sipA* mutations significantly reduced the enteropathogenicity of the corresponding mutant strains (Fig.4). The *copS* mutant was nearly as enteropathogenic as the wild-type strain (Fig.4). The *sopB, sopD* double mutant was significantly less enteropathogenic than the single *sopB* or *sipD* mutants indicating a synergistic effect of these mutations.

Finally the use of SPI-5-specific sequence to detect *Salmonella* serovars and distinguish them from other enteric bacterial pathogens was assessed by PCR. Three different sets of primers were used (see preceding explanation of fig.5). Primers PD1 and PD2 were derived from the *pipD* sequence, PB1 and PB2 - from *sopB*, and PA1 and PA2 - from *pipA*. Bacterial suspensions of individual strains were used as templates. The PCR reactions profile consisted of denaturation at 94°C for 5 min, followed by 25 cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 30 sec in a DNA thermal cycler. The analysis of PCR products by agarose gel electrophoresis showed that SPI-5-specific sequences of expected sizes were amplified when different *Salmonella* serotypes were used as template. No PCR products were detected when *Yersinia pseudotuberculosis*, *Shigella sonnei*, and enteropathogenic *E.coli* were used as templates (Fig.5A, B, and C). Thus, primers derived from SPI-5 are suitable for specific detection of *Salmonella* by PCR.

Cloning of the *sopA* Gene of *S.dublin* and Construction of *S.dublin* Mutants

Two pairs of degenerate oligonucleotide primers. A1(5-gaYWSNccNatHgaRttYgc-3'), and A1rev(5'-ggNaRNgcRaaYtcDatNgg-3'); A2(5'-gaRtggYtNggNccNgtNcaRga-3') and A2rev(5'-cYKtcYtgNacNggNccNaRccaYtc-3') were designed to match coding and non-coding DNA strands corresponding to amino acid sequences of two internal fragments of SopA obtained by the digestion of the protein with trypsin, D-S-P-I-E-F-A-L-P and E-W-L-G-P-V-Q-E-R, respectively. The capital letters in the oligonucleotide primer sequences above represent amino

acids. Thus where 'Y' appears it means that the oligonucleotide primer sequence may comprise any codon which codes for tyrosine; where 'W' appears it means that the oligonucleotide primer sequence may comprise any codon which codes for tryptophan etc. A1-A2rev and A1rev-A2 combinations of primers were used in PCR in an attempt to amplify a *sopA* gene fragment using *S.dublin* 2229 chromosomal DNA as template. The PCR reaction with A1-A2rev primers resulted in an approximately 1kb DNA fragment. This DNA fragment was cloned into pBluescript plasmid vector to yield pP80. The cloned DNA fragment was labelled and used as a probe in a Southern blot to detect *sopA*-specific fragments in different restriction enzyme digests of *S.dublin* chromosomal DNA. Two DNA fragments ~3kb EcoRI-*Bam*HI DNA fragment, and ~2.5kb *Pst*I-*Kpn*I DNA fragment that hybridised to the probe were identified and cloned into pBluescript to yield pMW80 and pMW79 respectively.

The *sopA* mutant of *S.dublin* was constructed as follows. An internal DNA fragment of *sopA* was amplified by PCR with the custom oligonucleotides SA1(5'-tgaagatatctcgaggcgcaattaat-3') and SA2(5'-taaggtgtttagatctttcggct-3') and was cloned into the suicide plasmid vector pDM4. The resulting plasmid was conjugated from *E.coli* S17.1 into *S.dublin* 2229 and Cml^r transconjugants were obtained. One of these clones was denoted *S.dublin* SA1 and chosen for further experiments. The correct insertion of the suicide vector into the *sopA* gene was confirmed by PCR.

SopA has a role in the induction of enteritis

To study the possible role of SopA in *Salmonella*-induced enteritis, we assessed the ability of the different *S.dublin* strains to induce fluid secretion and PMN influx in ligated ileal loops in two calves. Compared with the wild-type strain the *sopA* mutant *S.dublin* SA1 induced less fluid secretion and PMN influx (Fig 7) indicating that the SopA protein has a role in the induction of enteritis.

TABLE 1

Bacterial strains and plasmids used in this study.

Strain or Plasmid	Relevant genotype and description	Source of reference
E.coli		
DH5α	*recA*	BRL
S17-1λpir	RP4-2.Tc::Mu-Km::Tn7 (λpir)	(Simon et al., 1983)
S.dublin		
2229	wild-type, virulent field isolate	collection of IAH
B1	*sipB*::pB1 Cm^r	(Wood et al., 1996)
SB2	2229 with deletion of 1098 bp DNA fragment encoding an internal fragment of *sopB*	(Galyov et al., 1997)
PB2	2229 with deletion of 156 bp DNA fragment encoding an internal fragment of *pipB*	This study
PA1	*pipA*::pPIPA	This study
PD1	*pipD*::pPIPD	This study
KIN	*copS*::pKIN	This study

Plasmids		
pMW60	~3 kb *SalI-EcoRI sopB*-specific chromosomal DNA fragment cloned into pBluescript SK	(Galyov et al., 1997)
pMW61	Probe A positive, isolated from pBR322-based *Sau*IIIA library of *S.dublin* chromosomal DNA	This study
pMW59	Probe B positive ~3 kb *Hpa*I chromosomal DNA fragment cloned into pBluescript SK	This study
pMW58	Probe C positive ~2 kb *Eco*RI chromosomal DNA fragment cloned into pBluescript SK	This study
pMW57	Probe D positive ~3 kb *Cla*I-xxx chromosomal DNA fragment cloned into pBluescript SK	This study
pDM4	*oriR6K* Cm^r	(Milton et al., 1996)
pPIPA	PCR generated fragment of *pipA*	This study

	cloned into pDM4	
pPIBD	PCR generated fragment of *pipD* cloned into pDM4	This study
pKIN	PCR generated fragment of *copS* cloned into pDM4	This study
pDPIPB1	PCR generated DNA fragment spanning area covering the 5' and 3' parts of *pipB* and carrying the 156 bp long deletion of internal part of the gene cloned into pDM4.	This study

References

Allaoui, A., et al (1993) Infect Immun 61: 1707-1714.

Black, K., et al (1995) J Bacteriol 177: 6440-6448.

Blanc-Potard, A.-B., and Groisman, E.A. (1997) EMBO J 16: 5376-5385.

Eckmann, L., et al (1993) Infect Immun 61: 4569-4574.

Galán, J.E. (1996) Mol Microbiol 20: 263-271.

Galyov, E.E., et al (1997) Mol Microbiol 25: 903-912

Hayashi, T., et al (1993) Mol Microbiol 7: 657-667.

McCormick, B.A., et al(1993) J Cell Biol 123: 895-907.

McCormick, B.A., et al (1995a) J Cell Biol 131: 1599-1608.

McCormick, B.A., et al (1995b) Infect Immun 63: 2302-2309.

Mills, D.M., et al (1995) Mol Microbiol 15: 749-759.

Milton, D.L., et al (1996) J Bacteriol 178: 1310-1319.

Ochman, H., et al (1996) Proc Natl Acad Sci USA 93: 7800-7804.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989)

Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY 2nd Edn. Cold Spring Harbor Laboratory Press.

Shea, J.E., et al (1996) Proc Natl Acad Sci USA 93: 2593-2597.

Wallis, T.S., et al (1995) Infect Immun 63: 2755-2761.

Wood, M.W., et al (1996) Mol Microbiol 22: 327-338.

CLAIMS

1. An attenuated microorganism which is a mutant strain of *Salmonella* or another enteric bacteria and which, in comparison to the corresponding wild type strain, has reduced enteropathogenicity but substantially the same invasiveness, wherein when said mutant strain is *Salmonella*, it has one or more alterations in one or more genes selected from *sopA*, *sopD*, *sipA*, or a gene from within the pathogenicity island SPI-5, and wherein, where there is an alteration in the gene *sopB* of SPI-5, there is also an alteration in another of said genes, and wherein when said microorganism is a mutant strain of another enteric bacteria, there is an alteration in one or more homologous genes which correspond to a said *Salmonella* gene.

2. An attenuated microorganism according to claim 1 wherein, when when said mutant strain is Salmonella, it has alterations in a gene from within the pathogenicity island SPI-5 and said alterations are in one or more of the genes selected from *pipD*, *pipB and pipA*, and wherein when said microorganism is a mutant strain of another enteric bacteria, there is an alteration in one or more homologous genes which correspond to said *pipD*, *pipB or pipA* genes.

3. An attenuated microorganism according to claim 1 wherein, when said mutant strain is Salmonella and it has alterations in *sopB* and another of said genes, said other gene is *sopD*, and wherein when said microorganism is a mutant strain of another

enteric bacteria and has an alteration in a homologous gene which corresponds to sopB, there is also an alteration in a homologous gene which corresponds to sopD.

4. An attenuated microorganism strain according to any one of claims 1 to 3 which is in an isolated and/or purified form.

5. An attenuated microorganism strain according to any one of claims 1 to 4 which comprises an altered form of a serotype of *Salmonella* subspecies enterica.

6. An attenuated microorganism according to any one of claims 1 to 4 which is a mutant strain of *Shigella* or *E. coli*.

7. An attenuated microorganism according to any one of claims 1 to 6 which carries nucleotide sequence coding for an antigen different to that expressed by said attenuated microorganism.

8. A prophylactic or therapeutic medicament which comprises one or more attenuated microorganisms according to any one of claims 1 to 7.

9. A medicament according to claim 8 which is a vaccine.

10. A medicament according to claim 8 or claim 9 for use in the treatment of, or immunisation against, *Salmonella, Shigella or E.coli* infection.

11. A medicament according to any one of claims 8 to 10 which is formulated for administration by non-invasive or invasive injection or for mucosal administration.

12. A medicament according to any one of claims 8 to 11 which includes one or more pharmaceutically acceptable excipients, vehicles or carriers.

13. Use of an attenuated microorganism according to any one of claims 1 to 7 in the preparation of a medicament.

14. Use according to claim 13 wherein said medicament is one according to any one of claims 8 to 12.

15. A method of treatment or immunisation which comprises administering to a subject a medicament according to any one of claims 8 to 12.

16. A method according to claim 15 wherein the subject is a person or agricultural animal.

17. A method of treating, or immunising against, infection with *Salmonella, Shigella or E.coli*, comprising administering a medicament according to any one of claims 8 to 12.

18. A method of treating, or immunisation against, a medical condition, said method comprising administering a medicament which comprises an attenuated microorganism according to claim 7.

19. A substance which is a *sopA*, *pipD*, *orfX*, *pipC*, *pipB* or *pipA* polypeptide which has an amino acid sequence as shown in Figure 3 or Figure 6 or a peptide which comprises a characteristic part of a said *sopA*, *pipD*, *orfX*, *pipC*, *pipB* or *pipA* polypeptide.

20. A substance which is a polypeptide, mutant, variant, derivative, functional mimetic or allele of a said *sopA*, *pipD*, *orfX*, *pipC*, *pipB* or *pipA* polypeptide according to claim 19 or a peptide which comprises a characteristic part of said polypeptide which is a mutant, variant, derivative, functional mimetic or allele.

21. A substance which is a polypeptide having 20% or more sequence homology to a sopA, pipD, pipC, orfX, pipB or pipA polypeptide according to claim 19.

22. A substance which comprises a polypeptide or peptide according to any one of claims 19 to 21 and at least one other peptide sequence.

23. A nucleic acid molecule which codes for a substance being a polypeptide or peptide according to any one of claims 19 to 22.

24. A nucleic acid molecule according to claim 23 comprising, or complementary to, part or all of the *sopA*, *pipD*, *orfX*, *pipC*, *pipB* or *pipA* gene shown in Figure 3 or Figure 6.

25. A nucleic acid molecule according to claim 23 or claim 24 which further comprises a nucleotide

sequence functioning as a promoter or enhancer.

26. A nucleic acid molecule according to any one of claims 23 to 25 which is an oligonucleotide and can function as a primer in an amplification reaction.

27. A nucleic acid molecule according to claim 26 wherein said amplification reaction is for the detection of a *Salmonella*, *Shigella* or *E. coli*.

28. A vector comprising a nucleic acid molecule according to any one of claims 23 to 27.

29. A host cell transformed with a vector according to claim 28.

30. A method of identifying a gene which corresponds to a *Salmonella sopA*, *pipD*, *orfX*, *pipC*, *pipB* or *pipA* nucleic acid sequence which comprises using one or more nucleic acid molecules according to any one of claims 23 to 27 as a probe, the method comprising contacting the probe(s) and the test sample under hybridising conditions and observing whether hybridisation takes place.

31. A method according to claim 30 further including the step of isolating nucleic acid which has hybridised to the probe.

32. A method according to claim 30 or claim 31 for use in identifying the presence in a test sample of *Salmonella*, *Shigella* or *E. coli*.

33. An antibody or antibody fragment, derivative, functional equivalent or homologue of an antibody which is able to bind to a substance according to any one of claims 19 to 22 or a nucleic acid molecule according to any one of claims 23 to 27.

34. Use of a substance according to any one of claims 19 to 22 or of a nucleic acid molecule according to any one of claims 23 to 27 as an immunogen.

35. Use according to claim 34 in the production of monoclonal antibodies.

36. Use of an antibody or antibody fragment, derivative, functional equivalent or homologue of an antibody according to claim 33 in a method for determining the presence, amount or location in a sample of *Salmonella, Shigella or E.coli*.

37. A method of generating an attenuated microorganism which is a mutant strain of *Salmonella* for use as a vaccine or vaccine vector, which method comprises the steps of: (a) introducing into the microorganism a genomic mutation in one or more genes selected from *sopA, sopD, sipA,* or a gene from within the pathogenicity island SPI-5, wherein, where there is a mutation in the gene *sopB* of SPI-5, there is also a mutation in another of said genes; and (c) selecting and culturing a mutant microorganism which, in comparison to the corresponding wild type strain, has reduced enteropathogenicity but substantially the same invasiveness.

38. A method according to claim 37 wherein, when said mutant strain has alterations in *sopB* and another of said genes, said other gene is *sopD*.

39. Use of a substance according to any one of claims 19 to 22 or a nucleic acid molecule according to any one of claims 23 to 27 to screen for an agent which affects or modulates the activity of said substance.

40. A method of screening for an agent which modulates the activity of a substance of any one of claims 19 to 22, said method including the steps of: contacting one or more of said agents with said substance; testing the activity of the treated substance; and comparing the activity with that of the substance in a comparable reaction medium untreated with the test agent(s).

41. An agent as identified in a method according to claim 40.

42. Use of an agent according to claim 41 in the preparation of a medicament.

43. A prophylactic or therapeutic medicament which comprises one or more agents according to claim 41.

Figure 1

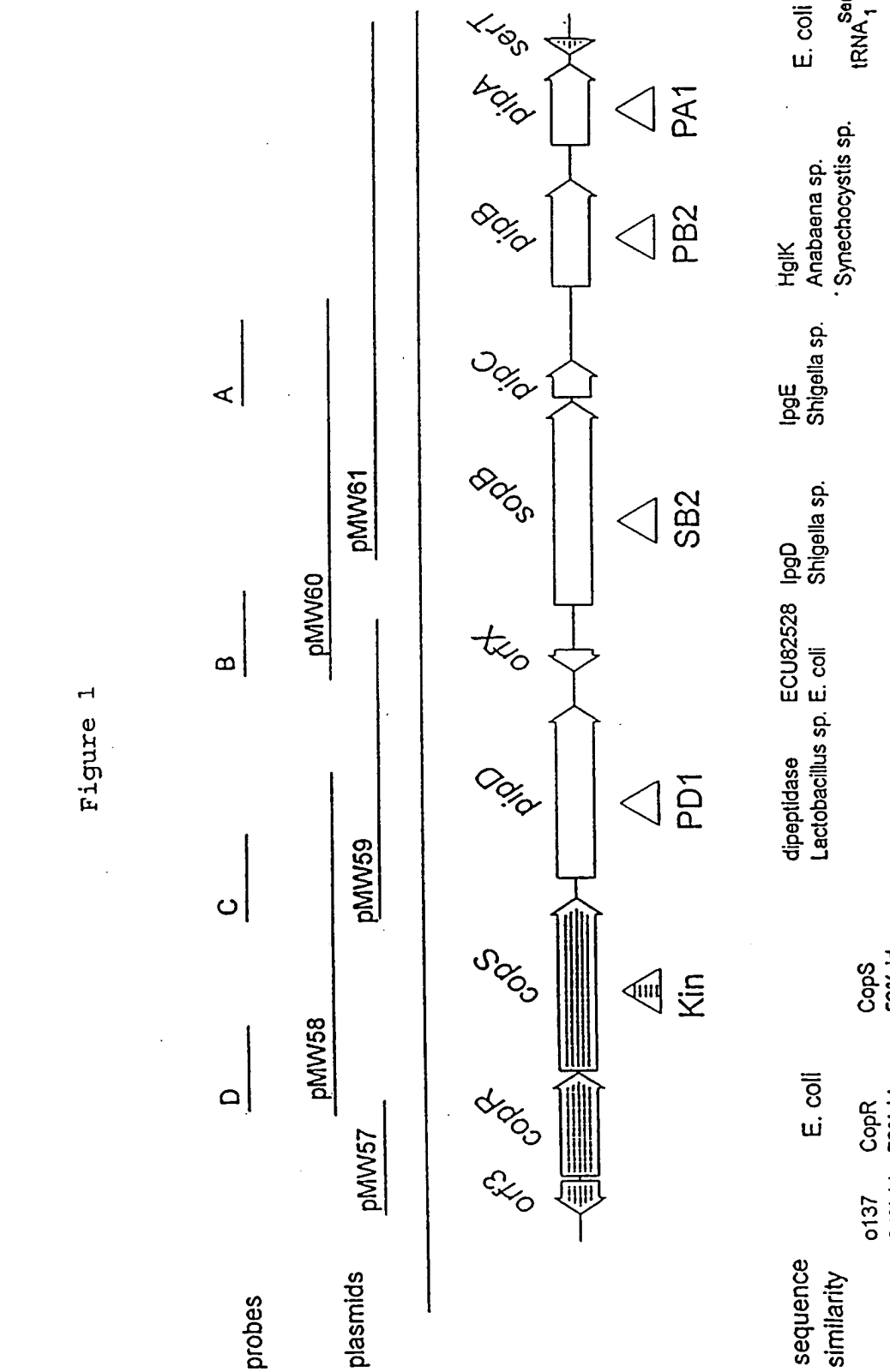

probes
D
C
B
A
plasmids
pMW57
pMW58
pMW59
pMW60
pMW61
orf3
copR
copS
Kin
pipD
PD1
orfX
sopB
SB2
pipC
pipB
PB2
pipA
PA1
sertRNA1Ser
sequence similarity
E. coli
o137 64% id
CopR 73% id
CopS 53% id
dipeptidase Lactobacillus sp.
ECU82528 E. coli
IpgD Shigella sp.
IpgE Shigella sp.
HglK Anabaena sp. Synechocystis sp.
E. coli

Figure 2

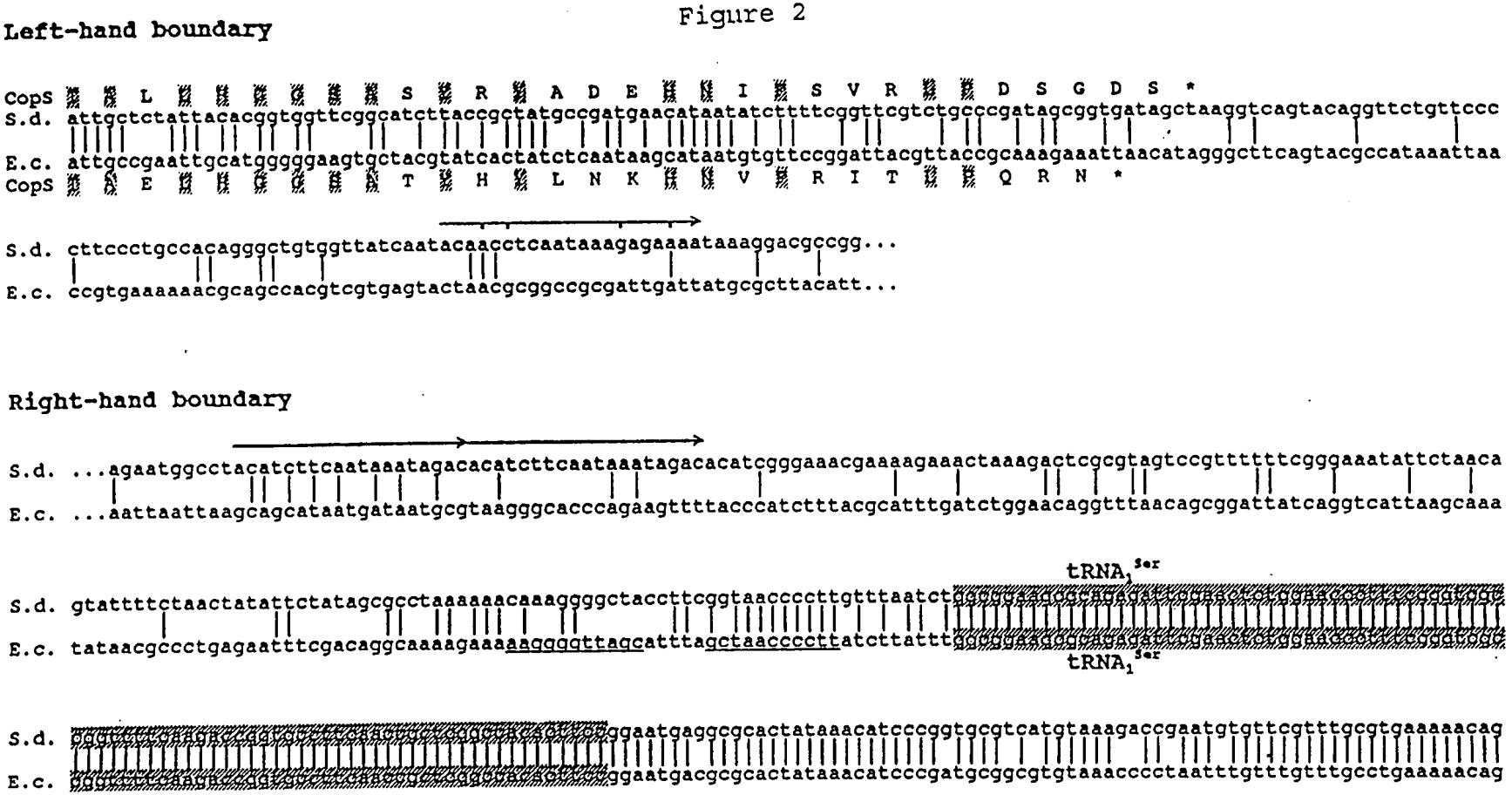

Left-hand boundary
CopS L S R A D E I S V R D S G D S *
S.d. attgctctattacacggtggttcggcatcttaccgctatgccgatgaacataatatcttttcggttcgtctgcccgatagcggtgatagctaaggtcagtacaggttctgttccc
E.c. attgccgaattgcatgggggaagtgctacgtatcactatctcaataagcataatgtgttccggattacgttaccgcaaagaaattaacatagggcttcagtacgccataaattaa
CopS E T H L N K V R I T Q R N *
S.d. cttccctgccacagggctgtggttatcaatacaacctcaataaagagaaaataaaggacgccgg...
E.c. ccgtgaaaaaacgcagccacgtcgtgagtactaacgcggccgcgattgattatgcgcttacatt...
Right-hand boundary
S.d. ...agaatggcctacatcttcaataaatagacacatcttcaataaatagacacatcgggaaacgaaaagaaactaaagactcgcgtagtccgttttttcgggaaatattctaaca
E.c. ...aattaattaagcagcataatgataatgcgtaagggcacccagaagttttacccatctttacgcatttgatctggaacaggtttaacagcggattatcaggtcattaagcaaa
tRNA1Ser
S.d. gtattttctaactatattctatagcgcctaaaaaacaaaggggctaccttcggtaaccccttgtttaatct
E.c. tataacgccctgagaatttcgacaggcaaaagaaaaaggggttagcatttagctaacccccttatcttatttg
tRNA1Ser
S.d.
E.c. ggaatgacgcgcactataaacatcccgatgcggcgtgtaaaccctaatttgtttgtttgcctgaaaaacag

Fig. 2

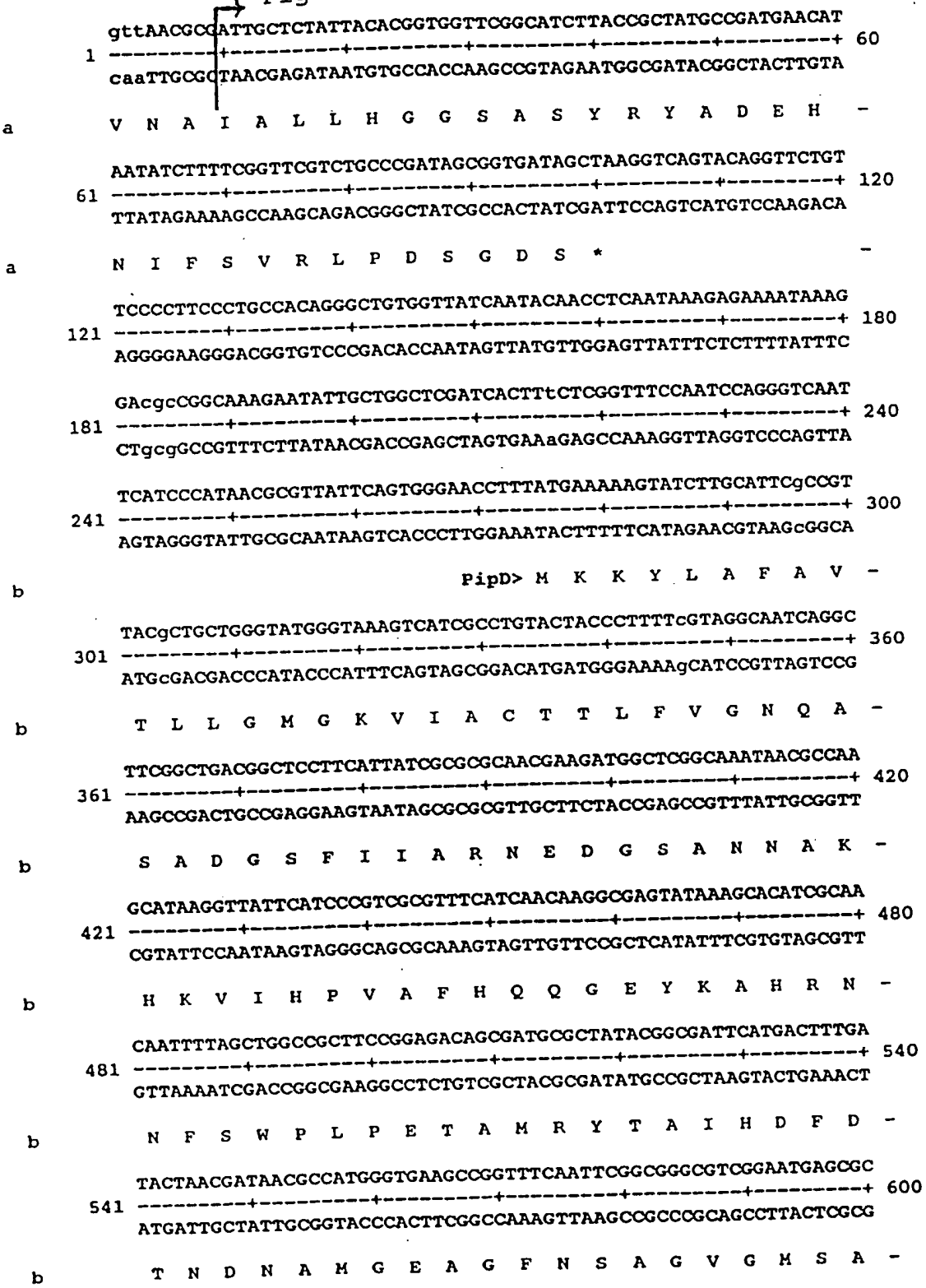


Figure 3

Figure 3 cont. (i)

```
     AACGGAAACCATTTATAACGGCAGAGCGGCGCTGGCTGCCGATCCTTACGTGACAAAAAC
601  ---------+---------+---------+---------+---------+---------+ 660
     TTGCCTTTGGTAAATATTGCCGTCTCGCCGCGACCGACGGCTAGGAATGCACTGTTTTTG

b      T  E  T  I  Y  N  G  R  A  A  L  A  A  D  P  Y  V  T  K  T  -

     AGGAATCACGGAAGACGCCATTGAGTCCGTGATCCTGCCAGTGGCGCAATCGGCGCGTCA
661  ---------+---------+---------+---------+---------+---------+ 720
     TCCTTAGTGCCTTCTGCGGTAACTCAGGCACTAGGACGGTCACCGCGTTAGCCGCGCAGT

b      G  I  T  E  D  A  I  E  S  V  I  L  P  V  A  Q  S  A  R  Q  -

     GGGCGCCAAATTACTGGGAGACATTATTGAACAAAAAGGCGCGGGCGAAGGTTTCGGCGT
721  ---------+---------+---------+---------+---------+---------+ 780
     CCCGCGGTTTAATGACCCTCTGTAATAACTTGTTTTTCCGCGCCCGCTTCCAAAGCCGCA

b      G  A  K  L  L  G  D  I  I  E  Q  K  G  A  G  E  G  F  G  V  -

     CGCGTTTATTGATAGCAAAGAGATATGGTATCTGGAGACGGGAAGCGGTCATCAATGGCT
781  ---------+---------+---------+---------+---------+---------+ 840
     GCGCAAATAACTATCGTTTCTCTATACCATAGACCTCTGCCCTTCGCCAGTAGTTACCGA

b      A  F  I  D  S  K  E  I  W  Y  L  E  T  G  S  G  H  Q  W  L  -

     GGCAGTACGACTTCCGGCAGATAGCTATTTCGTTTCCGCCAATCAGGGACGTTTACGCCA
841  ---------+---------+---------+---------+---------+---------+ 900
     CCGTCATGCTGAAGGCCGTCTATCGATAAAGCAAAGGCGGTTAGTCCCTGCAAATGCGGT

b      A  V  R  L  P  A  D  S  Y  F  V  S  A  N  Q  G  R  L  R  H  -

     TTACGATCCGAATGATAACGCGAATTATATGGCGTCACCAACGTTAGTAAGCTTTGCGAA
901  ---------+---------+---------+---------+---------+---------+ 960
     AATGCTAGGCTTACTATTGCGCTTAATATACCGCAGTGGTTGCAATCATTCGAAACGCTT

b      Y  D  P  N  D  N  A  N  Y  M  A  S  P  T  L  V  S  F  A  K  -

     AAAGCAGGGATTATATGATCCGGCCCGCGGCGAATTCGACTTTCATCAAGCCTATTCACA
961  ---------+---------+---------+---------+---------+---------+ 1020
     TTTCGTCCCTAATATACTAGGCCGGGCGCCGCTTAAGCTGAAAGTAGTTCGGATAAGTGT

b      K  Q  G  L  Y  D  P  A  R  G  E  F  D  F  H  Q  A  Y  S  Q  -

     GGATAACAAAAACGATACCACCTATAATTATCCGCGCGTCTGGACGCTACAACACCAGTT
1021 ---------+---------+---------+---------+---------+---------+ 1080
     CCTATTGTTTTTGCTATGGTGGATATTAATAGGCGCGCAGACCTGCGATGTTGTGGTCAA

b      D  N  K  N  D  T  T  Y  N  Y  P  R  V  W  T  L  Q  H  Q  F  -

     TAATCCGCATCTGGATACGGTCGTTAGCCCAGGGGAAACATTTCCCGTTTTTTTAACGCC
1081 ---------+---------+---------+---------+---------+---------+ 1140
     ATTAGGCGTAGACCTATGCCAGCAATCGGGTCCCCTTTGTAAAGGGCAAAAAAATTGCGG

b      N  P  H  L  D  T  V  V  S  P  G  E  T  F  P  V  F  L  T  P  -
```

Figure 3 cont. (ii)

```
     AATAACGAAGATCAGCGTGGCCGCAGTAAAAAACGCGTTACGCAATCACTATCAGGGAAC
1141 ---------+---------+---------+---------+---------+---------+ 1200
     TTATTGCTTCTAGTCGCACCGGCGTCATTTTTTGCGCAATGCGTTAGTGATAGTCCCTTG

b      I  T  K  I  S  V  A  A  V  K  N  A  L  R  N  H  Y  Q  G  T  -

     GTCGCACGACCCTTATGCCAGTCATAATCCACAAGAACCATGGCGACCTATATCCGTTTT
1201 ---------+---------+---------+---------+---------+---------+ 1260
     CAGCGTGCTGGGAATACGGTCAGTATTAGGTGTTCTTGGTACCGCTGGATATAGGCAAAA

b      S  H  D  P  Y  A  S  H  N  P  Q  E  P  W  R  P  I  S  V  F  -

     TCGTACCCAGGAGTCACATATTTTACAGGTCAGACCGAAATTACCGCAGGCTATCGGCAA
1261 ---------+---------+---------+---------+---------+---------+ 1320
     AGCATGGGTCCTCAGTGTATAAAATGTCCAGTCTGGCTTTAATGGCGTCCGATAGCCGTT

b      R  T  Q  E  S  H  I  L  Q  V  R  P  K  L  P  Q  A  I  G  N  -

     CGTAGAATACATCGCCTATGGAATGCCATCTCTTAGCGTCTATCTCCCCTATTATCAGGG
1321 ---------+---------+---------+---------+---------+---------+ 1380
     GCATCTTATGTAGCGGATACCTTACGGTAGAGAATCGCAGATAGAGGGGATAATAGTCCC

b      V  E  Y  I  A  Y  G  M  P  S  L  S  V  Y  L  P  Y  Y  Q  G  -

     GATGCGTCATTATCAACCCGGAGATGATAAAGGACCGATCGGGCGAGCAACGACTCTACC
1381 ---------+---------+---------+---------+---------+---------+ 1440
     CTACGCAGTAATAGTTGGGCCTCTACTATTTCCTGGCTAGCCCGCTCGTTGCTGAGATGG

b      M  R  H  Y  Q  P  G  D  D  K  G  P  I  G  R  A  T  T  L  P  -

     TACTGGACATTCCGCACGCTGCAAACGCTGGTTATGCAGGACTACAATGCGTTTGCGCCA
1441 ---------+---------+---------+---------+---------+---------+ 1500
     ATGACCTGTAAGGCGTGCGACGTTTGCGACCAATACGTCCTGATGTTACGCAAACGCGGT

b      T  G  H  S  A  R  C  K  R  W  L  C  R  T  T  M  R  L  R  Q  -

     GATGTGCAACACGCCTGCAAAACATTTGAACAGCAAACAGCTAAGCAGCAGtATAAGATG
1501 ---------+---------+---------+---------+---------+---------+ 1560
     CTACACGTTGTGCGGACGTTTTGTAAACTTGTCGTTTGTCGATTCGTCGTCaTATTCTAC

b      M  C  N  T  P  A  K  H  L  N  S  K  Q  L  S  S  S  I  R  W  -

     GAGCAGAGCTATCTGAGaTtATATGCGTCGCATCCGAAAGAAGCGCAACGCTTACTGCAA
1561 ---------+---------+---------+---------+---------+---------+ 1620
     CTCGTCTCGATAGACTCtAaTATACGCAGCGTAGGCTTTCTTCGCGTTGCGAATGACGTT

b      S  R  A  I  *

     AATTTTGAAGATAAAACGATGCAAAATGCGCAGACGCTCGCCCGTCGCCTGACCAATAAT
1621 ---------+---------+---------+---------+---------+---------+ 1680
     TTAAAACTTCTATTTTGCTACGTTTTACGCGTCTGCGAGCGGGCAGCGGACTGGTTATTA

     ATTACTACGACAATGACTTACCGTACAGATATGAAATATCACTTTTCAAGTACGgAACCA
1681 ---------+---------+---------+---------+---------+---------+ 1740
     TAATGATGCTGTTACTGAATGGCATGTCTATACTTTATAGTGAAAAGTTCATGCcTTGGT
```

Figure 3 cont. (iii)

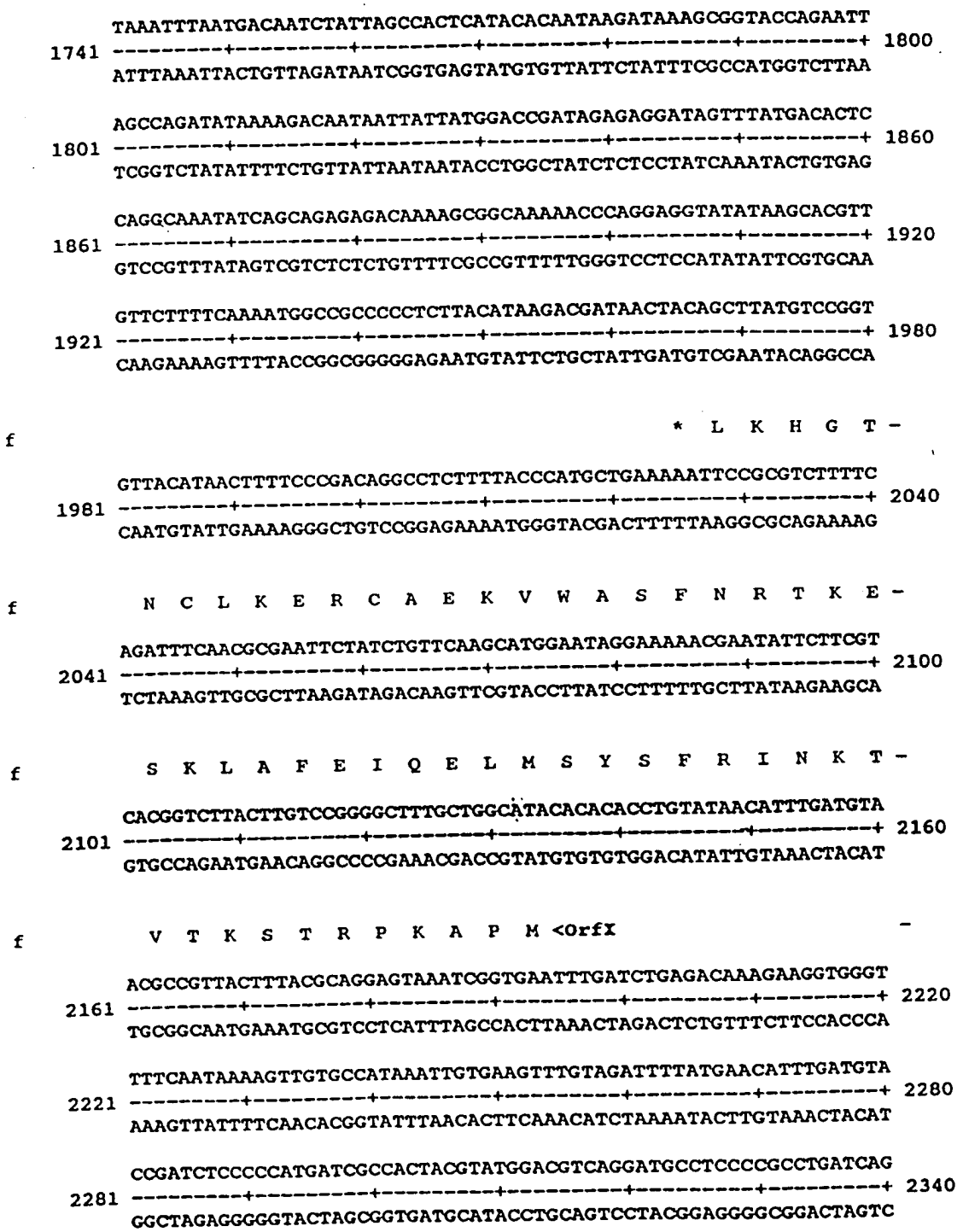


Figure 3 cont. (iv)

```
     AAGCGTTTCCTCATTAAAAAGGACATTTTTTAAAGTTCCTAGTGCATAAAAGTCACATCC
2341 ---------+---------+---------+---------+---------+---------+ 2400
     TTCGCAAAGGAGTAATTTTTCCTGTAAAAAATTTCAAGGATCACGTATTTTCAGTGTAGG

     TTTTAAAGGGTTGTTAACCCTGTTGAATGTTCCCACTCCCCTATTCAGGAATATTAAAAA
2401 ---------+---------+---------+---------+---------+---------+ 2460
     AAAATTTCCCAACAATTGGGACAACTTACAAGGGTGAGGGGATAAGTCCTTATAATTTTT

     CGCTATGCAAATACAGAGCTTCTATCACTCAGCTTCACTAAAAACCCAGGAGGCTTTTAA
2461 ---------+---------+---------+---------+---------+---------+ 2520
     GCGATACGTTTATGTCTCGAAGATAGTGAGTCGAAGTGATTTTTGGGTCCTCCGAAAATT

b      SopB> M  Q  I  Q  S  F  Y  H  S  A  S  L  K  T  Q  E  A  F  K  -

     AAGCCTACAAAAAACCTTATACAACGGAATGCAGATTCTCTCAGGCCAGGGCAAAGCGCC
2521 ---------+---------+---------+---------+---------+---------+ 2580
     TTCGGATGTTTTTTGGAATATGTTGCCTTACGTCTAAGAGAGTCCGGTCCCGTTTCGCGG

b       S  L  Q  K  T  L  Y  N  G  M  Q  I  L  S  G  Q  G  K  A  P  -

     GGCTAAAGCGCCCGACGCTCGCCCGGAAATTATTGTCCTGCGAGAACCTGGCGCGACATG
2581 ---------+---------+---------+---------+---------+---------+ 2640
     CCGATTTCGCGGGCTGCGAGCGGGCCTTTAATAACAGGACGCTCTTGGACCGCGCTGTAC

b       A  K  A  P  D  A  R  P  E  I  I  V  L  R  E  P  G  A  T  W  -

     GGGGAATTATCTACAGCATCAGAAGACGTCTAACCACTCGCTGCATAACCTCTATAACTT
2641 ---------+---------+---------+---------+---------+---------+ 2700
     CCCCTTAATAGATGTCGTAGTCTTCTGCAGATTGGTGAGCGACGTATTGGAGATATTGAA

b       G  N  Y  L  Q  H  Q  K  T  S  N  H  S  L  H  N  L  Y  N  L  -

     ACGGCGCGATCTTCTTACGGTCGGGGCAACCGTTCTGGGTAAACAAGACCCGGTTCTAAC
2701 ---------+---------+---------+---------+---------+---------+ 2760
     TGCCGCGCTAGAAGAATGCCAGCCCCGTTGGCAAGACCCATTTGTTCTGGGCCAAGATTG

b       R  R  D  L  L  T  V  G  A  T  V  L  G  K  Q  D  P  V  L  T  -

     GTCAATGGCAAACCAAATGGAGTTAGCCAAAGTTAAAGCGGACCGGCCAGCAACAAAACA
2761 ---------+---------+---------+---------+---------+---------+ 2820
     CAGTTACCGTTTGGTTTACCTCAATCGGTTTCAATTTCGCCTGGCCGGTCGTTGTTTTGT

b       S  M  A  N  Q  M  E  L  A  K  V  K  A  D  R  P  A  T  K  Q  -

     AGAAGAAGCCGCGGCAAAAGCATTGAAGAAAAATCTTATCGAACTTATTGCAGCACGCAC
2821 ---------+---------+---------+---------+---------+---------+ 2880
     TCTTCTTCGGCGCCGTTTTCGTAACTTCTTTTTAGAATAGCTTGAATAACGTCGTGCGTG

b       E  E  A  A  A  K  A  L  K  K  N  L  I  E  L  I  A  A  R  T  -

     TCAGCAGCAGGATGGCTTACCTGCAAAAGAAGCTCATCGCTTTGCGGCAGTAGCGTTTAG
2881 ---------+---------+---------+---------+---------+---------+ 2940
     AGTCGTCGTCCTACCGAATGGACGTTTTCTTCGAGTAGCGAAACGCCGTCATCGCAAATC

b       Q  Q  Q  D  G  L  P  A  K  E  A  H  R  F  A  A  V  A  F  R  -
```

Figure 3 cont. (v)

```
      AGACGCTCAGGACAAGCAGCTTAATAACCAGCCCTGGCAAACCATAAAAAATACACTCAC
2941  ---------+---------+---------+---------+---------+---------+ 3000
      TCTGCGAGTCCTGTTCGTCGAATTATTGGTCGGGACCGTTTGGTATTTTTTATGTGAGTG

b      D  A  Q  D  K  Q  L  N  N  Q  P  W  Q  T  I  K  N  T  L  T  -

      GCATAACGGGCATCACTATACCAACACGCAGCTCCCTGCCGCAGAGATGAAAATCGGCGC
3001  ---------+---------+---------+---------+---------+---------+ 3060
      CGTATTGCCCGTAGTGATATGGTTGTGCGTCGAGGGACGGCGTCTCTACTTTTAGCCGCG

b      H  N  G  H  H  Y  T  N  T  Q  L  P  A  A  E  M  K  I  G  A  -

      AAAAGATATCTTTCCCAGTGCTTATGAGGGAAAGGGCGTATGCAGTTGGGATACCAAGAA
3061  ---------+---------+---------+---------+---------+---------+ 3120
      TTTTCTATAGAAAGGGTCACGAATACTCCCTTTCCCGCATACGTCAACCCTATGGTTCTT

b      K  D  I  F  P  S  A  Y  E  G  K  G  V  C  S  W  D  T  K  N  -

      TATTCATCACGCCAATAATTTGTGGATGTCCACGGTGAGTGTGCATGAGGACGGTAAAGA
3121  ---------+---------+---------+---------+---------+---------+ 3180
      ATAAGTAGTGCGGTTATTAAACACCTACAGGTGCCACTCACACGTACTCCTGCCATTTCT

b      I  H  H  A  N  N  L  W  M  S  T  V  S  V  H  E  D  G  K  D  -

      TAAAACGCTTTTCTGCGGGATACGTCATGGCGTGCTTTCCCCCTATCATGAAAAAGATCC
3181  ---------+---------+---------+---------+---------+---------+ 3240
      ATTTTGCGAAAAGACGCCCTATGCAGTACCGCACGAAAGGGGGATAGTACTTTTTCTAGG

b      K  T  L  F  C  G  I  R  H  G  V  L  S  P  Y  H  E  K  D  P  -

      GCTTCTGCGTCAGGTCGGCGCTGAAAACAAAGCCAAAGAAGTATTAACTGCGGCACTTTT
3241  ---------+---------+---------+---------+---------+---------+ 3300
      CGAAGACGCAGTCCAGCCGCGACTTTTGTTTCGGTTTCTTCATAATTGACGCCGTGAAAA

b      L  L  R  Q  V  G  A  E  N  K  A  K  E  V  L  T  A  A  L  F  -

      TAGTAAACCTGAGTTGCTTAACAAAGCCTTAGCGGGCGAGGCGGTAAGCCTGAAACTGGT
3301  ---------+---------+---------+---------+---------+---------+ 3360
      ATCATTTGGACTCAACGAATTGTTTCGGAATCGCCCGCTCCGCCATTCGGACTTTGACCA

b      S  K  P  E  L  L  N  K  A  L  A  G  E  A  V  S  L  K  L  V  -

      ATCCGTGGGGTTACTCACCGCGTCGAATATTTTCGGCAAAGAGGGAACGATGGTCGAGGA
3361  ---------+---------+---------+---------+---------+---------+ 3420
      TAGGCACCCCAATGAGTGGCGCAGCTTATAAAAGCCGTTTCTCCCTTGCTACCAGCTCCT

b      S  V  G  L  L  T  A  S  N  I  F  G  K  E  G  T  M  V  E  D  -

      TCAAATGCGCGCATGGCAATCGTTGACCCAGCCGGGAAAAATGATTCATTTAAAAATCCG
3421  ---------+---------+---------+---------+---------+---------+ 3480
      AGTTTACGCGCGTACCGTTAGCAACTGGGTCGGCCCTTTTTACTAAGTAAATTTTTAGGC

b      Q  M  R  A  W  Q  S  L  T  Q  P  G  K  M  I  H  L  K  I  R  -
```

Figure 3 cont. (vi)

```
      CAATAAAGATGGCGATCTACAGACGGTAAAAATAAAACCGGACGTCGCCGCATTTAATGT
3481  ---------+---------+---------+---------+---------+---------+ 3540
      GTTATTTCTACCGCTAGATGTCTGCCATTTTTATTTTGGCCTGCAGCGGCGTAAATTACA

b       N  K  D  G  D  L  Q  T  V  K  I  K  P  D  V  A  A  F  N  V  -

      GGGTGTTAATGAGTTGGCGCTCAAGCTCGGCTTTGGCCTTAAGGCATCGGATAGCTATAA
3541  ---------+---------+---------+---------+---------+---------+ 3600
      CCCACAATTACTCAACCGCGAGTTCGAGCCGAAACCGGAATTCCGTAGCCTATCGATATT

b       G  V  N  E  L  A  L  K  L  G  F  G  L  K  A  S  D  S  Y  N  -

      TGCCGAGGCGCTATATCAGTTATTAGGCAATGATTTACGCCCTGAAGCCAGACCAGGTGG
3601  ---------+---------+---------+---------+---------+---------+ 3660
      ACGGCTCCGCGATATAGTCAATAATCCGTTACTAAATGCGGGACTTCGGTCTGGTCCACC

b       A  E  A  L  Y  Q  L  L  G  N  D  L  R  P  E  A  R  P  G  G  -

      CTGGGTTGGCGAATGGCTGGCACAATACCCGGATAATTATGAGGTCGTCAATACATTAGC
3661  ---------+---------+---------+---------+---------+---------+ 3720
      GACCCAACCGCTTACCGACCGTGTTATGGGCCTATTAATACTCCAGCAGTTATGTAATCG

b       W  V  G  E  W  L  A  Q  Y  P  D  N  Y  E  V  V  N  T  L  A  -

      GCGCCAGATTAAGGATATATGGAAAAATAACCAACATCATAAAGATGGCGGCGAACCCTA
3721  ---------+---------+---------+---------+---------+---------+ 3780
      CGCGGTCTAATTCCTATATACCTTTTTATTGGTTGTAGTATTTCTACCGCCGCTTGGGAT

b       R  Q  I  K  D  I  W  K  N  N  Q  H  H  K  D  G  G  E  P  Y  -

      TAAACTCGCACAACGCCTTGCCATGTTAGCCCATGAAATTGACGCGGTACCCGCCTGGAA
3781  ---------+---------+---------+---------+---------+---------+ 3840
      ATTTGAGCGTGTTGCGGAACGGTACAATCGGGTACTTTAACTGCGCCATGGGCGGACCTT

b       K  L  A  Q  R  L  A  M  L  A  H  E  I  D  A  V  P  A  W  N  -

      TTGTAAAAGCGGCAAAGATCGTACAGGGATGATGGATTCAGAAATCAAGCGAGAGATCAT
3841  ---------+---------+---------+---------+---------+---------+ 3900
      AACATTTTCGCCGTTTCTAGCATGTCCCTACTACCTAAGTCTTTAGTTCGCTCTCTAGTA

b       C  K  S  G  K  D  R  T  G  M  M  D  S  E  I  K  R  E  I  I  -

      TTCCTTACATCAGACCCATATGTTAAGTGCGCCTGGCAGTCTTCCGGATAGCGGTGGACA
3901  ---------+---------+---------+---------+---------+---------+ 3960
      AAGGAATGTAGTCTGGGTATACAATTCACGCGGACCGTCAGAAGGCCTATCGCCACCTGT

b       S  L  H  Q  T  H  M  L  S  A  P  G  S  L  P  D  S  G  G  Q  -

      GAAAATTTTCCAAAAAGTATTACTGAATAGCGGTAACCCTGGAGATTCAGAACCAAATAC
3961  ---------+---------+---------+---------+---------+---------+ 4020
      CTTTTAAAAGGTTTTTCATAATGACTTATCGCCATTGGGACCTCTAAGTCTTGGTTTATG

b       K  I  F  Q  K  V  L  L  N  S  G  N  P  G  D  S  E  P  N  T  -
```

Figure 3 cont. (vii)

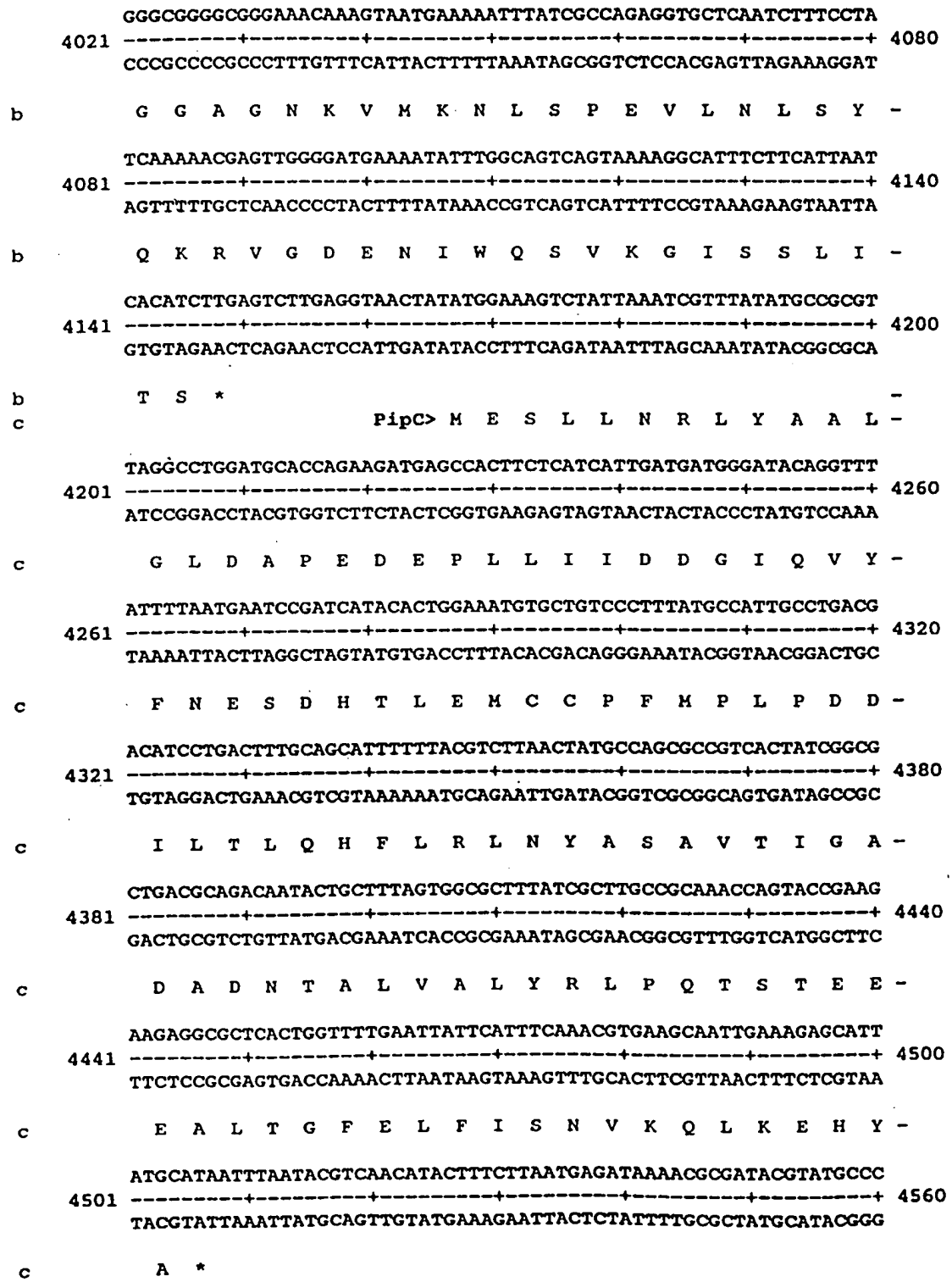


Figure 3 cont. (viii)

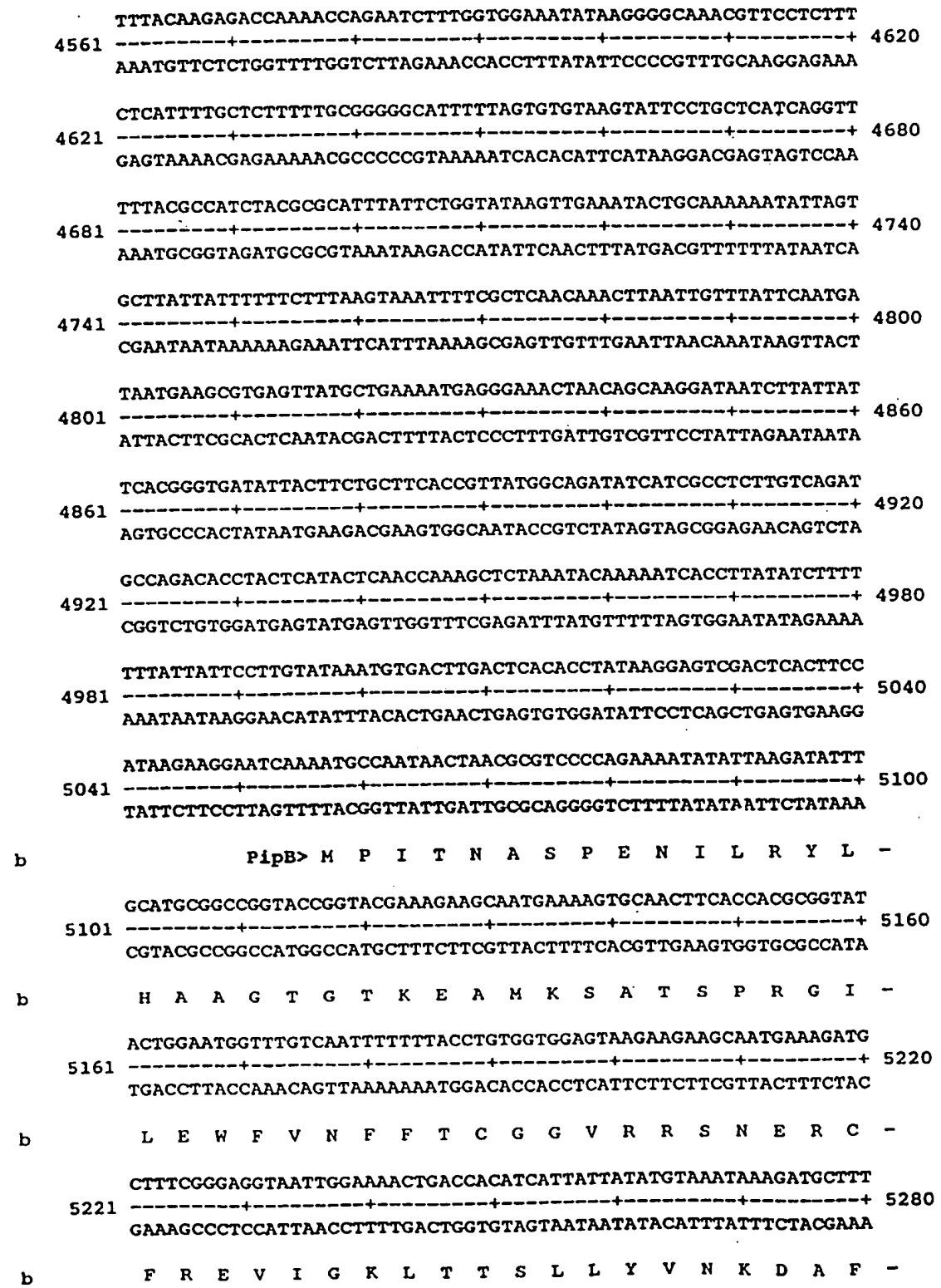


Figure 3 cont. (ix)

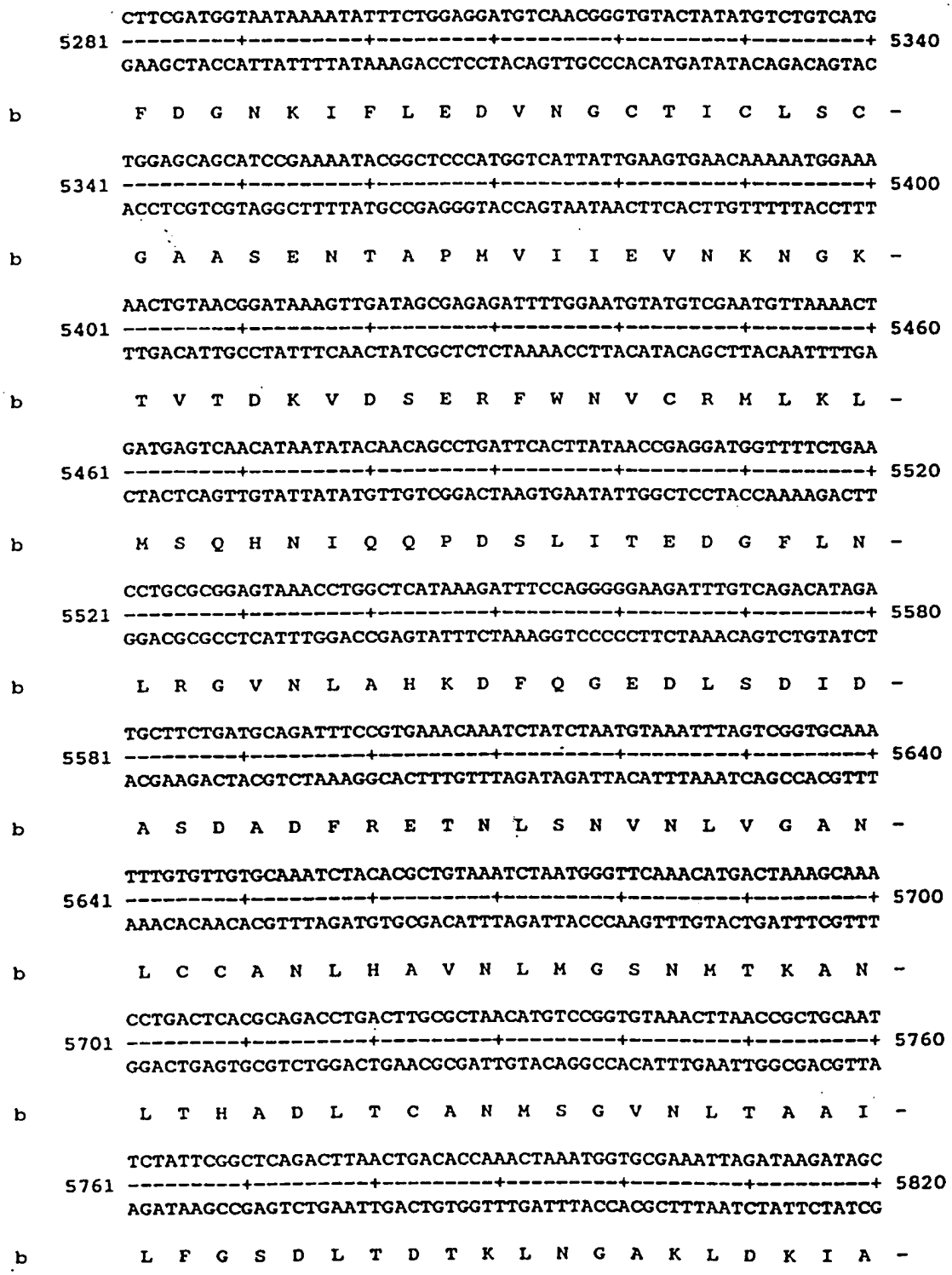


Figure 3 cont. (x)

```
      TCTAACTTTAGCGAAAGCATTAACAGGAGCCGATCTGACAGGTAGTCAACATACCCCTAC
5821  ---------+---------+---------+---------+---------+---------+ 5880
      AGATTGAAATCGCTTTCGTAATTGTCCTCGGCTAGACTGTCCATCAGTTGTATGGGGATG

b      L  T  L  A  K  A  L  T  G  A  D  L  T  G  S  Q  H  T  P  T  -

      TCCACTCCCGGATTACAATGATAGAACTCTTTCCCCCCATCCGATATTTTAGTCGAGATA
5881  ---------+---------+---------+---------+---------+---------+ 5940
      AGGTGAGGGCCTAATGTTACTATCTTGAGAAAGGGGGGTAGGCTATAAAATCAGCTCTAT

b      R  L  P  D  Y  N  D  R  T  L  S  P  H  P  I  F  *           -

      AAGGGATTTTATAAACAAGAAGTATTCAAACAGAGGCCCCCTTGTTTTATTAAATAAACC
5941  ---------+---------+---------+---------+---------+---------+ 6000
      TTCCCTAAAATATTTGTTCTTCATAAGTTTGTCTCCGGGGGAACAAAATAATTTATTTGG

      CCGCCCCTAAGTTTCATTATAAATAACATTTTCAGCGTATTACTTGTTGGCTTTGTCTGA
6001  ---------+---------+---------+---------+---------+---------+ 6060
      GGCGGGGATTCAAAGTAATATTTATTGTAAAAGTCGCATAATGAACAACCGAAACAGACT

      ATCATAGCGTTATCTGTATGTGGCACATTAACTAAAAACACTATTATTGTTTAATTTAAA
6061  ---------+---------+---------+---------+---------+---------+ 6120
      TAGTATCGCAATAGACATACACCGTGTAATTGATTTTTGTGATAATAACAAATTAAATTT

      TAATTCATAATTGTAGTCAGGAAATAAGAAGTTATGCTTCCGGTCACCTACAGATTAATA
6121  ---------+---------+---------+---------+---------+---------+ 6180
      ATTAAGTATTAACATCAGTCCTTTATTCTTCAATACGAAGGCCAGTGGATGTCTAATTAT

a                                    PipA> M  L  P  V  T  Y  R  L  I  -

      CCTCAAAGCGGAGTATCCACATATGGATTAAATACCGCAGATACACCTGTTTTCCCCGAT
6181  ---------+---------+---------+---------+---------+---------+ 6240
      GGAGTTTCGCCTCATAGGTGTATACCTAATTTATGGCGTCTATGTGGACAAAAGGGGCTA

a      P  Q  S  G  V  S  T  Y  G  L  N  T  A  D  T  P  V  F  P  D  -

      ATTCCCGAACATGCACCAAACCCCTCCATGCTACGCCTTGCTCATGACAGCCTTGCCATA
6241  ---------+---------+---------+---------+---------+---------+ 6300
      TAAGGGCTTGTACGTGGTTTGGGGAGGTACGATGCGGAACGAGTACTGTCGGAACGGTAT

a      I  P  E  H  A  P  N  P  S  M  L  R  L  A  H  D  S  L  A  I  -

      AACAGTGAATTCCGTCTGGAGCCAGAGTGTGTGGTGGAGTACCTTATCTCAGGCGCGGGT
6301  ---------+---------+---------+---------+---------+---------+ 6360
      TTGTCACTTAAGGCAGACCTCGGTCTCACACACCACCTCATGGAATAGAGTCCGCGCCCA

a      N  S  E  F  R  L  E  P  E  C  V  V  E  Y  L  I  S  G  A  G  -

      GGAATAGACCCTGATACAGAAATTGATGACGACACTTATAACGAATGCTACGATGAACTA
6361  ---------+---------+---------+---------+---------+---------+ 6420
      CCTTATCTGGGACTATGTCTTTAACTACTGCTGTGAATATTGCTTACGATGCTACTTGAT

a      G  I  D  P  D  T  E  I  D  D  D  T  Y  N  E  C  Y  D  E  L  -
```

Figure 3 cont. (xi)

```
      TCCTCCGTACTTCAAAATGCGTATACCCAAAGCGAAACATTCCGCAGACTGATGAATTAC
6421  ---------+---------+---------+---------+---------+---------+ 6480
      AGGAGGCATGAAGTTTTACGCATATGGGTTTCGCTTTGTAAGGCGTCTGACTACTTAATG

a       S  S  V  L  Q  N  A  Y  T  Q  S  E  T  F  R  R  L  M  N  Y  -

      GCATATGAAAAAGAACTACATGATGTGGAGCAGCGCTGGCTACTGGGGGCAGGCGAAGCC
6481  ---------+---------+---------+---------+---------+---------+ 6540
      CGTATACTTTTTCTTGATGTACTACACCTCGTCGCGACCGATGACCCCCGTCCGCTTCGG

a       A  Y  E  K  E  L  H  D  V  E  Q  R  W  L  L  G  A  G  E  A  -

      TTTGAAACTTCCGTGGCTCAGGAACACTTCAAACTTTCAGAAGGCAGGAAAGTTATTTGT
6541  ---------+---------+---------+---------+---------+---------+ 6600
      AAACTTTGAAGGCACCGAGTCCTTGTGAAGTTTGAAAGTCTTCCGTCCTTTCAATAAACA

a       F  E  T  S  V  A  Q  E  H  F  K  L  S  E  G  R  K  V  I  C  -

      CTCAATCTGGACGATTCTGATGATTCATATACCGAACATTATGAAAGTAACGAAGGACCA
6601  ---------+---------+---------+---------+---------+---------+ 6660
      GAGTTAGACCTGCTAAGACTACTAAGTATATGGCTTGTAATACTTTCATTGCTTCCTGGT

a       L  N  L  D  D  S  D  D  S  Y  T  E  H  Y  E  S  N  E  G  P  -

      CAACTTTTTGACACAAAACGTTCATTTATTCATGAAGTTGTACATGCACTGACCCATCTT
6661  ---------+---------+---------+---------+---------+---------+ 6720
      GTTGAAAAACTGTGTTTTGCAAGTAAATAAGTACTTCAACATGTACGTGACTGGGTAGAA

a       Q  L  F  D  T  K  R  S  F  I  H  E  V  V  H  A  L  T  H  L  -

      CAGGATAAAGAAGAAAATCATCCAAGAGGCCCTGTTGTCGAATATACCAACATTATTCTG
6721  ---------+---------+---------+---------+---------+---------+ 6780
      GTCCTATTTCTTCTTTTAGTAGGTTCTCCGGGACAACAGCTTATATGGTTGTAATAAGAC

a       Q  D  K  E  E  N  H  P  R  G  P  V  V  E  Y  T  N  I  I  L  -

      AAAGAGATGGGGCATCCTTCACCTCCCAGAATggccTACATCTTCAATAAATAGACACAT
6781  ---------+---------+---------+---------+---------+---------+ 6840
      TTTCTCTACCCCGTAGGAAGTGGAGGGTCTTAccggATGTAGAAGTTATTTATCTGTGTA

a       K  E  M  G  H  P  S  P  P  R  M  A  Y  I  F  N  K  *        -

      CTTCAATAAATAGACACATCGGGAAACGAAAAGAAACTAAAGACTCGCGTAGTCCGTTTT
6841  ---------+---------+---------+---------+---------+---------+ 6900
      GAAGTTATTTATCTGTGTAGCCCTTTGCTTTTCTTTGATTTCTGAGCGCATCAGGCAAAA

      TTCGGGAAATATTCTAACAGTATTTTCTAACTATATTCTATAGCGCCTAAAAAACAAAGG
6901  ---------+---------+---------+---------+---------+---------+ 6960
      AAGCCCTTTATAAGATTGTCATAAAAGATTGATATAAGATATCGCGGATTTTTTGTTTCC

      GGCTACCTTCGGTAACCCCTTGTTTAATCTGGCGGAAGCGCAGAGATTCGAACTCTGGAA
6961  ---------+---------+---------+---------+---------+---------+ 7020
      CCGATGGAAGCCATTGGGGAACAAATTAGACCGCCTTCGCGTCTCTAAGCTTGAGACCTT
```

Figure 3 cont. (xii)

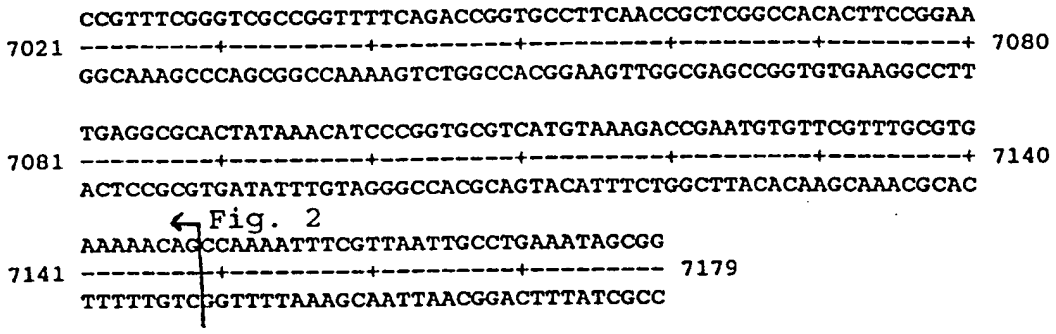


Figure 4

Effect of mutations in genes of SPI-5
on Salmonella-induced enteropathogenesis

Figure 4a

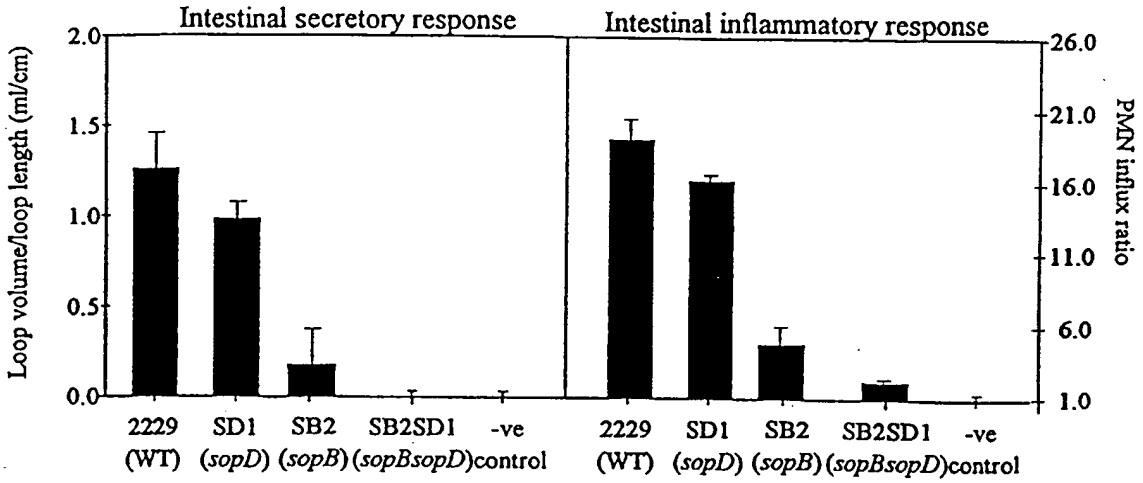


Figure 4b

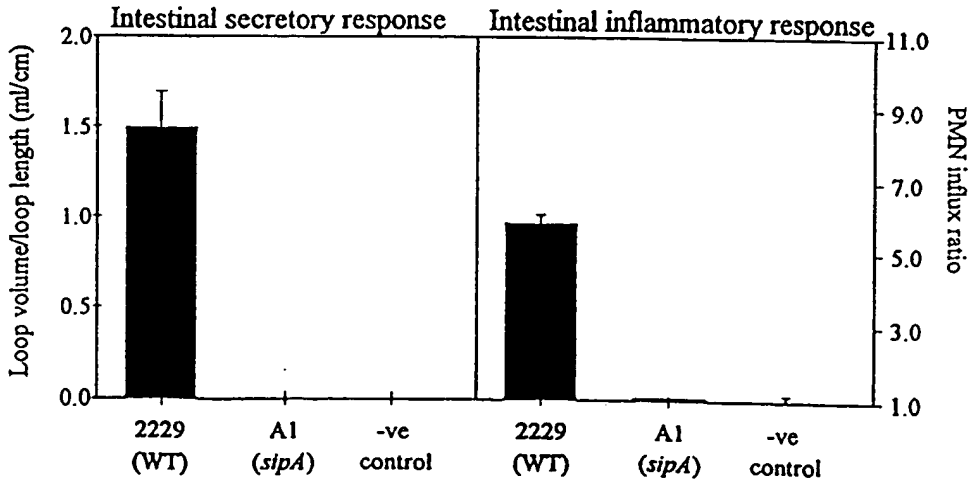


Figure 4c

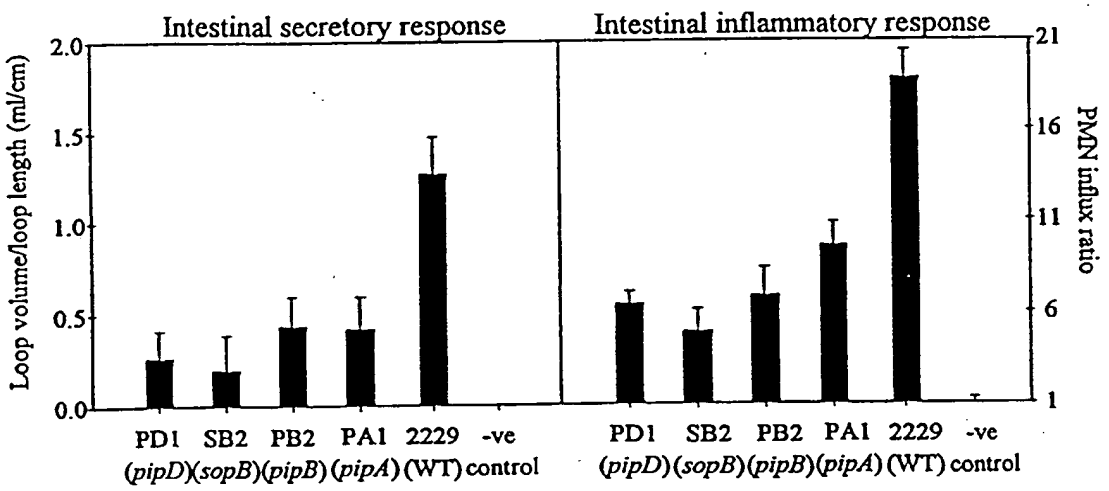

Intestinal secretory response
Intestinal inflammatory response
Loop volume/loop length (ml/cm)
PMN influx ratio
2.0
1.5
1.0
0.5
0.0
21
16
11
6
1
PD1 SB2 PB2 PA1 2229 -ve
(pipD) (sopB) (pipB) (pipA) (WT) control
PD1 SB2 PB2 PA1 2229 -ve
(pipD) (sopB) (pipB) (pipA) (WT) control

Figure 5

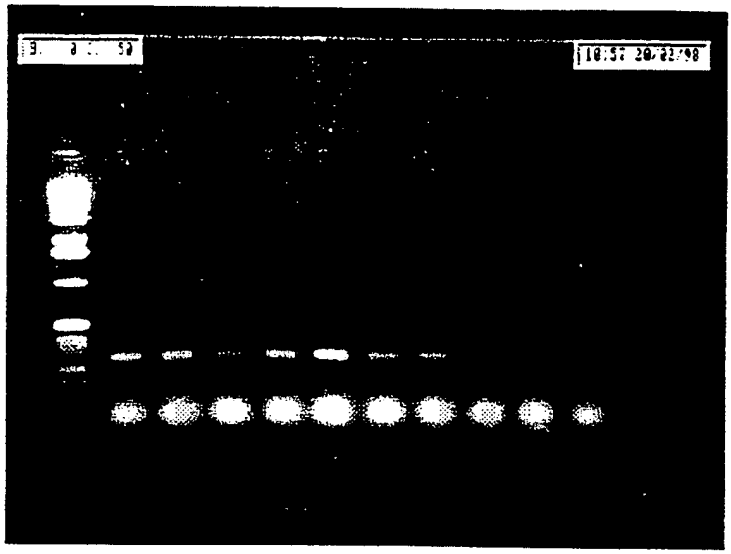


Figure 5 cont. (i)

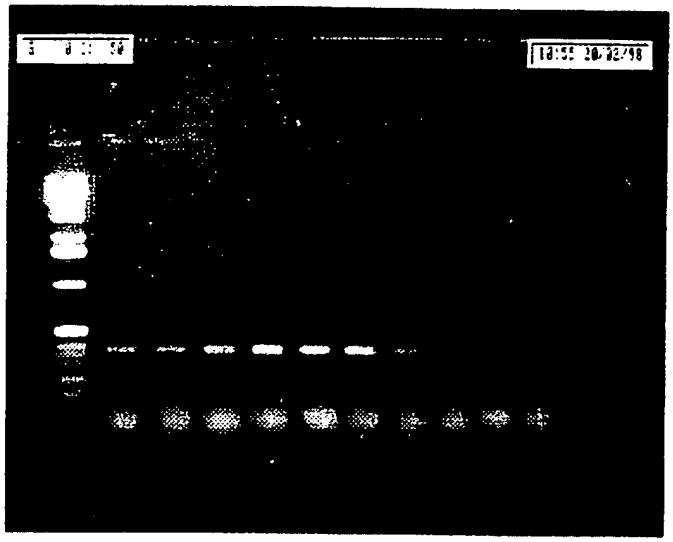


Figure 5 cont. (ii)

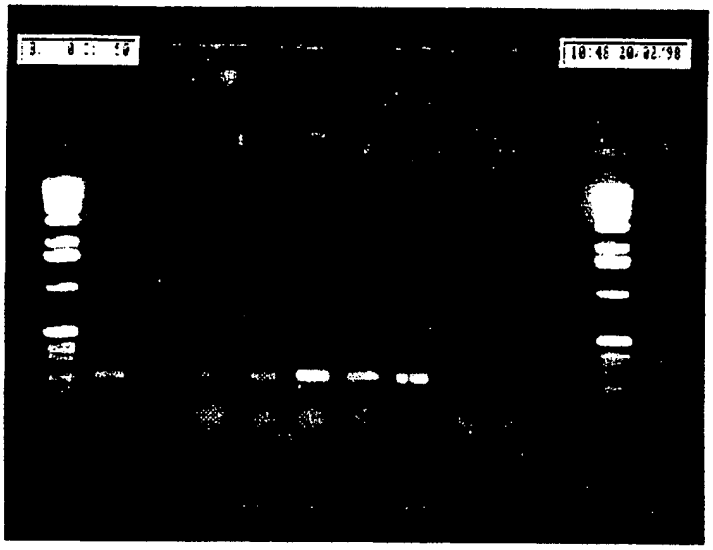


Figure 6

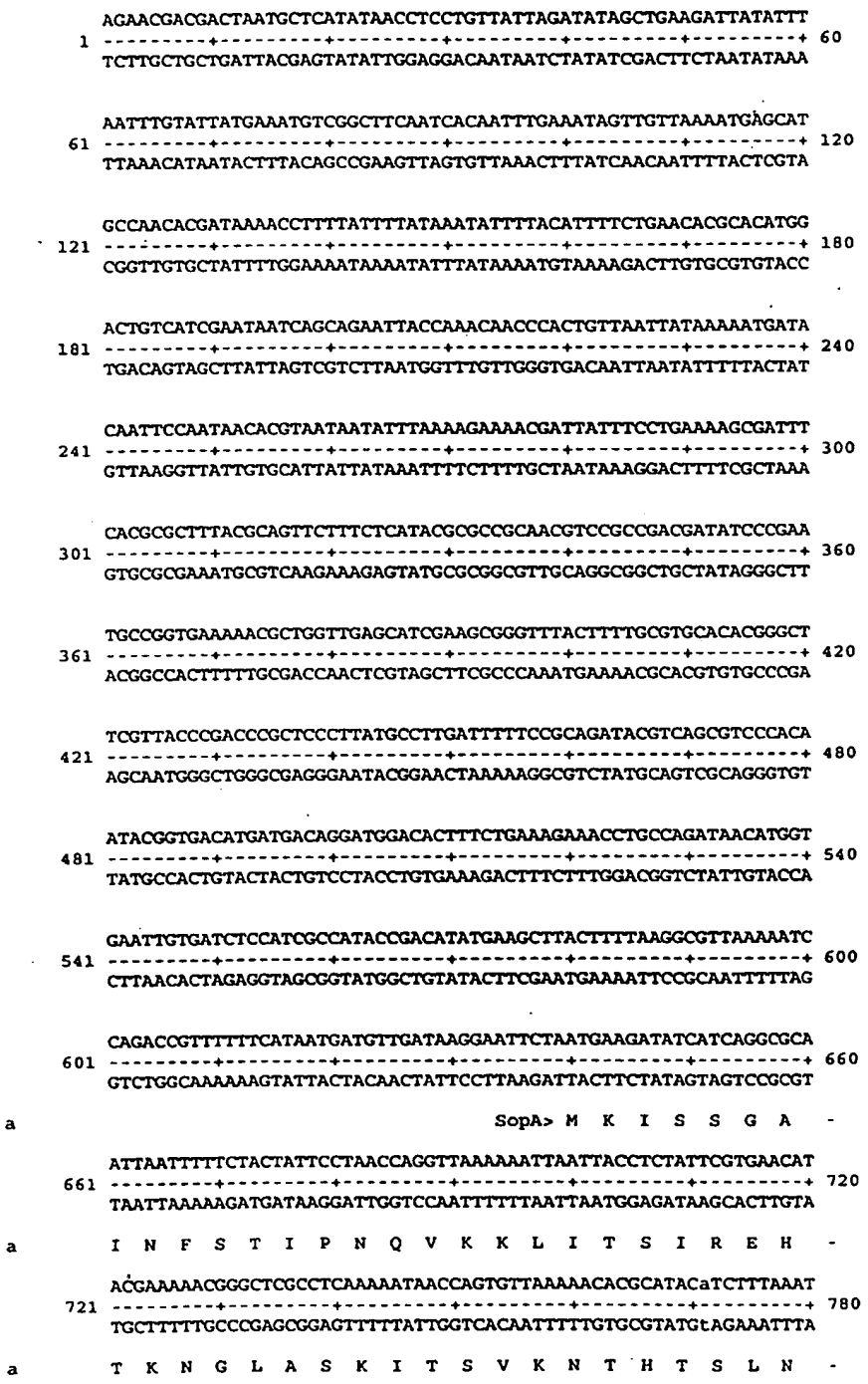

AGAACGACGACTAATGCTCATATAACCTCCTGTTATTAGATATAGCTGAAGATTATATTT
1
TCTTGCTGCTGATTACGAGTATATTGGAGGACAATAATCTATATCGACTTCTAATATAAA 60
AATTTGTATTATGAAATGTCGGCTTCAATCACAATTTGAAATAGTTGTTAAAATGAGCAT
61
TTAAACATAATACTTTACAGCCGAAGTTAGTGTTAAACTTTATCAACAATTTTACTCGTA 120
GCCAACACGATAAAACCTTTTATTTTATAAATATTTTACATTTTCTGAACACGCACATGG
121
CGGTTGTGCTATTTTGGAAAATAAAATATTTATAAAATGTAAAAGACTTGTGCGTGTACC 180
ACTGTCATCGAATAATCAGCAGAATTACCAAACAACCCACTGTTAATTATAAAAATGATA
181
TGACAGTAGCTTATTAGTCGTCTTAATGGTTTGTTGGGTGACAATTAATATTTTTACTAT 240
CAATTCCAATAACACGTAATAATATTTAAAAGAAAACGATTATTTCCTGAAAAGCGATTT
241
GTTAAGGTTATTGTGCATTATTATAAATTTTCTTTTGCTAATAAAGGACTTTTCGCTAAA 300
CACGCGCTTTACGCAGTTCTTTCTCATACGCGCCGCAACGTCCGCCGACGATATCCCGAA
301
GTGCGCGAAATGCGTCAAGAAAGAGTATGCGCGGCGTTGCAGGCGGCTGCTATAGGGCTT 360
TGCCGGTGAAAAACGCTGGTTGAGCATCGAAGCGGGTTTACTTTTGCGTGCACACGGGCT
361
ACGGCCACTTTTTGCGACCAACTCGTAGCTTCGCCCAAATGAAAACGCACGTGTGCCCGA 420
TCGTTACCCGACCCGCTCCCTTATGCCTTGATTTTTCCGCAGATACGTCAGCGTCCCACA
421
AGCAATGGGCTGGGCGAGGGAATACGGAACTAAAAAGGCGTCTATGCAGTCGCAGGGTGT 480
ATACGGTGACATGATGACAGGATGGACACTTTCTGAAAGAAACCTGCCAGATAACATGGT
481
TATGCCACTGTACTACTGTCCTACCTGTGAAAGACTTTCTTTGGACGGTCTATTGTACCA 540
GAATTGTGATCTCCATCGCCATACCGACATATGAAGCTTACTTTTAAGGCGTTAAAAATC
541
CTTAACACTAGAGGTAGCGGTATGGCTGTATACTTCGAATGAAAATTCCGCAATTTTTAG 600
CAGACCGTTTTTTCATAATGATGTTGATAAGGAATTCTAATGAAGATATCATCAGGCGCA
601
GTCTGGCAAAAAAGTATTACTACAACTATTCCTTAAGATTACTTCTATAGTAGTCCGCGT 660
a SopA> M K I S S G A -
ATTAATTTTTCTACTATTCCTAACCAGGTTAAAAAATTAATTACCTCTATTCGTGAACAT
661
TAATTAAAAAGATGATAAGGATTGGTCCAATTTTTTAATTAATGGAGATAAGCACTTGTA 720
a I N F S T I P N Q V K K L I T S I R E H -
ACGAAAAACGGGCTCGCCTCAAAAATAACCAGTGTTAAAAACACGCATACaTCTTTAAAT
721
TGCTTTTTGCCCGAGCGGAGTTTTTATTGGTCACAATTTTTGTGCGTATGtAGAAATTTA 780
a T K N G L A S K I T S V K N T H T S L N -

Figure 6 cont. (i)

```
      GAAAAATTTAAAACAGGAAAGGACTCACCGATTGAGTTCGCGTTACCACAAAAAATAAAA
  781 ---------+---------+---------+---------+---------+---------+ 840
      CTTTTTAAATTTTGTCCTTTCCTGAGTGGCTAACTCAAGCGCAATGGTGTTTTTTATTTT

a         E  K  F  K  T  G  K  D  S  P  I  E  F  A  L  P  Q  K  I  K  -

      GACTTCTTTCAGCCGAAAGATAAAAACACCTTAAACAAAACATTGATTACTGTTAAAAAT
  841 ---------+---------+---------+---------+---------+---------+ 900
      CTGAAGAAAGTCGGCTTTCTATTTTTGTGGAATTTGTTTTGTAACTAATGACAATTTTTA

a         D  F  F  Q  P  K  D  K  N  T  L  N  K  T  L  I  T  V  K  N  -

      ATTAAAGATACAAATAATGCAGGCAAGAAAAATATTTCAGCAGAAGATGTCTCAAAAATG
  901 ---------+---------+---------+---------+---------+---------+ 960
      TAATTTCTATGTTTATTACGTCCGTTCTTTTTATAAAGTCGTCTTCTACAGAGTTTTTAC

a         I  K  D  T  N  N  A  G  K  K  N  I  S  A  E  D  V  S  K  M  -

      AATGCAGCATTCATGCGTAAGCATATTGCAAATCAAACATGTGATTATAATTACAGAATG
  961 ---------+---------+---------+---------+---------+---------+ 1020
      TTACGTCGTAAGTACGCATTCGTATAACGTTTAGTTTGTACACTAATATTAATGTCTTAC

a         N  A  A  F  M  R  K  H  I  A  N  Q  T  C  D  Y  N  Y  R  M  -

      ACAGGTGCGGCCCCGCTCCCCGGTGGAGTCTCTGTATCAGCCAATAACAGGCCCACGGTT
 1021 ---------+---------+---------+---------+---------+---------+ 1080
      TGTCCACGCCGGGGCGAGGGGCCACCTCAGAGACATAGTCGGTTATTGTCCGGGTGCCAA

a         T  G  A  A  P  L  P  G  G  V  S  V  S  A  N  N  R  P  T  V  -

      TCTGAAGGTAGAACACCACCAGTATCCCCCTCCCTCTCACTTCAGGCTACGTCTTCCCCG
 1081 ---------+---------+---------+---------+---------+---------+ 1140
      AGACTTCCATCTTGTGGTGGTCATAGGGGGAGGGAGAGTGAAGTCCGATGCAGAAGGGGC

a         S  E  G  R  T  P  P  V  S  P  S  L  S  L  Q  A  T  S  S  P  -

      TCATCACCTGCCGACTGGGCTAAGAAACTCACAGATGCAGTTTTACGACAGAAAGCCGGA
 1141 ---------+---------+---------+---------+---------+---------+ 1200
      AGTAGTGGACGGCTGACCCGATTCTTTGAGTGTCTACGTCAAAATGCTGTCTTTCGGCCT

a         S  S  P  A  D  W  A  K  K  L  T  D  A  V  L  R  Q  K  A  G  -

      GAAACCCTTACGGCCGCAGATCGCGATTTTTCAAACGCAGATTTCCGTAATATTACATTC
 1201 ---------+---------+---------+---------+---------+---------+ 1260
      CTTTGGGAATGCCGGCGTCTAGCGCTAAAAAGTTTGCGTCTAAAGGCATTATAATGTAAG

a         E  T  L  T  A  A  D  R  D  F  S  N  A  D  F  R  N  I  T  F  -

      AGCAAAATATTGCCCCCCAGCTTCATGGAGCGAGACGGCGATATTATTAAGGGGTTCAAC
 1261 ---------+---------+---------+---------+---------+---------+ 1320
      TCGTTTTATAACGGGGGGTCGAAGTACCTCGCTCTGCCGCTATAATAATTCCCCAAGTTG

a         S  K  I  L  P  P  S  F  M  E  R  D  G  D  I  I  K  G  F  N  -

      TTTTCAAATTCAAAATTTACTTATTCTGATATATCTCATTTACATTTTGACGAATGCCGA
 1321 ---------+---------+---------+---------+---------+---------+ 1380
      AAAAGTTTAAGTTTTAAATGAATAAGACTATATAGAGTAAATGTAAAACTGCTTACGGCT

a         F  S  N  S  K  F  T  Y  S  D  I  S  H  L  H  F  D  E  C  R  -

      TTCACTTATTCGACACTGAGTGATGTAGTCTGCAGTAATACGAAATTTAGTAATTCAGAC
 1381 ---------+---------+---------+---------+---------+---------+ 1440
      AAGTGAATAAGCTGTGACTCACTACATCAGACGTCATTATGCTTTAAATCATTAAGTCTG

a         F  T  Y  S  T  L  S  D  V  V  C  S  N  T  K  F  S  N  S  D  -
```

Figure 6 cont. (ii)

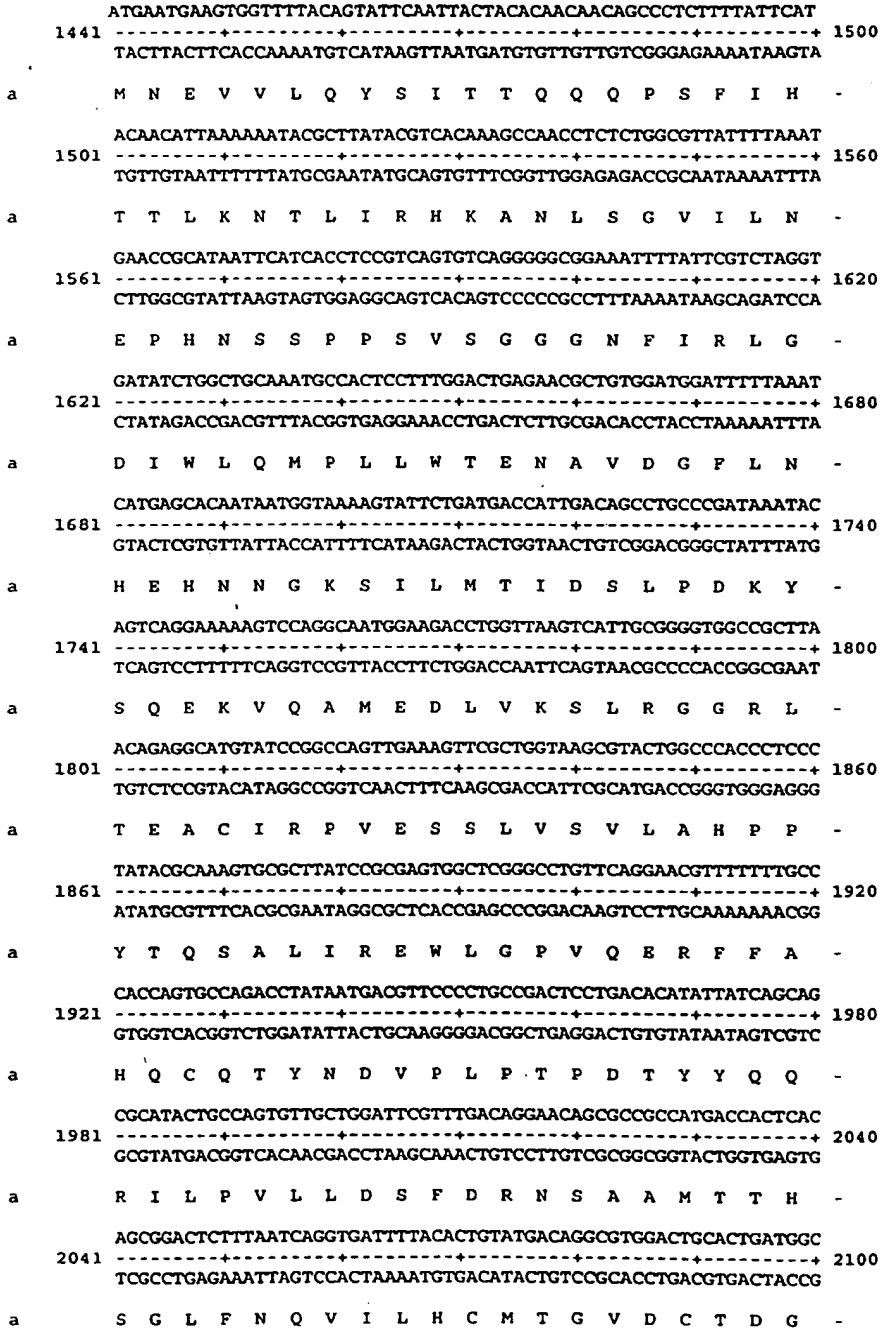


Figure 6 cont. (iii)

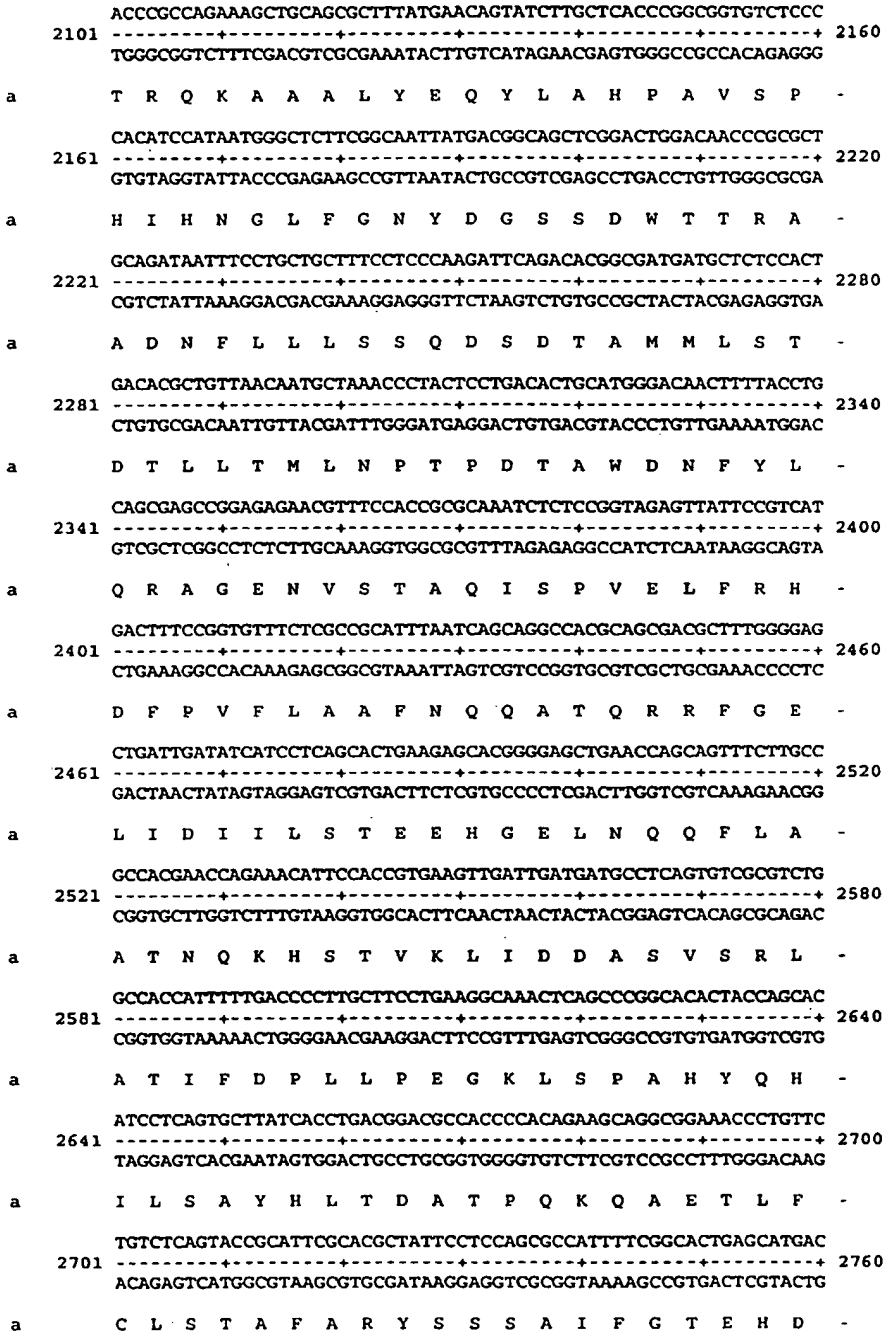


Figure 6 cont. (iv)

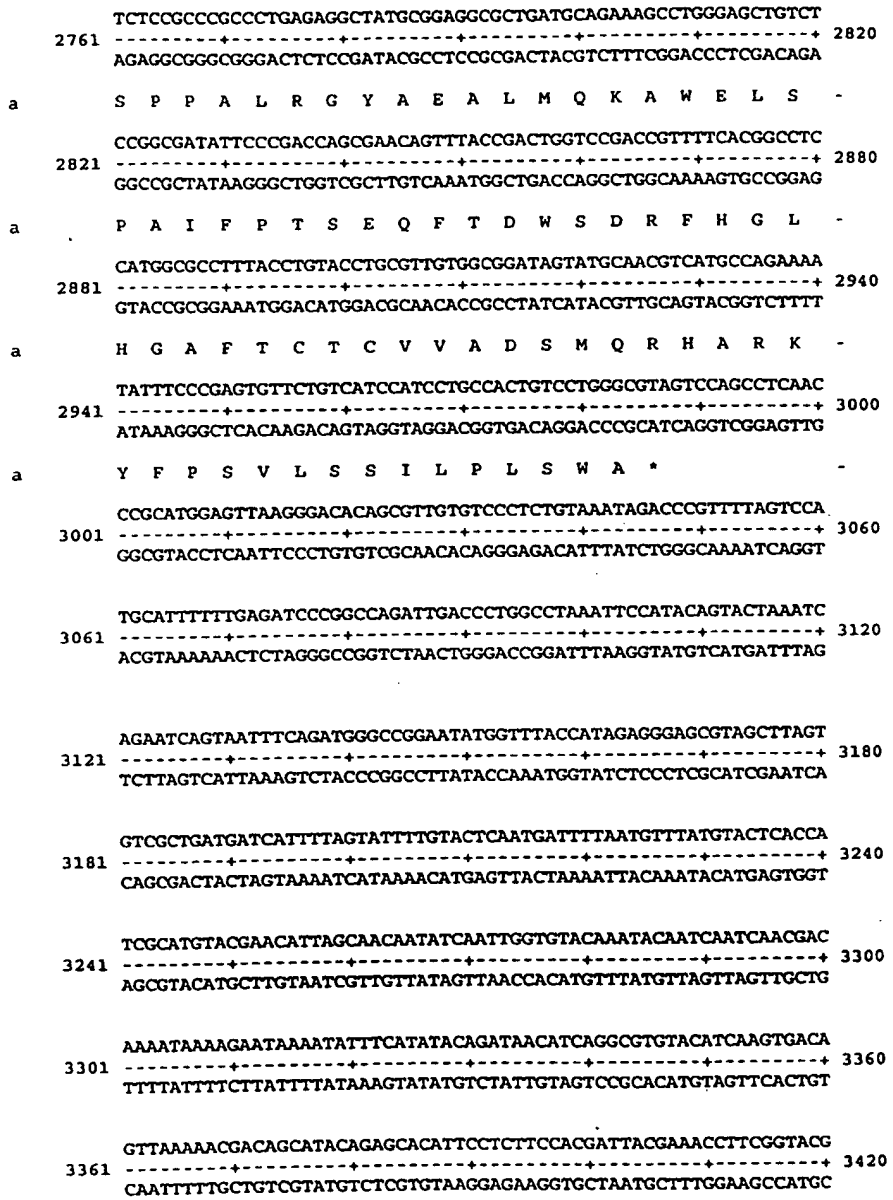


Figure 7

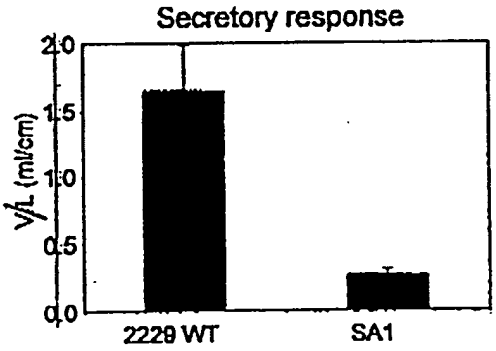


Figure 7 cont.